The Biometrics Section of the
American Statistical Association

BIOMETRICS

TABLE OF CONTENTS

Estimation of Bacterial Densities by Means of the "Most Probable Number" . . . . . W. G. COCHRAN 105

Statistical Analysis in Sanitary Engineering Laboratory Studies ROLF ELIASSEN 117

The Application of Statistical Techniques to Sewage Treatment Processes . . . . . J. W. FERTIG and A. N. HELLER 127

Fiducial Intervals for Variance Components . . IRWIN BROSS 136

Determining Scales and the Use of Transformations in Studies on Weight Per Locule of Tomato Fruit . . . LEROY POWERS 145

Queries 164

Abstracts 169

The Biometric Society 195

Proceedings of the Second International Biometric Conference Continued

Recent Applications of Biometrical Methods in Genetics

(1) Experimental Techniques in Plant Improvement FRANK YATES 200

(2) The Analysis of Selection Curves . LUIGI L. CAVALLI 208

(3) Scores for the Estimation of Parameters D. J. FINNEY 221

Industrial Applications of Biometry

Biometric Methods in Chemical Industry . . O. L. DAVIES 228

News and Notes 231

Number 2 June 1950 Volume 6

Material for *Biometrics* should be addressed to the Chairman of the Editorial Board, Institute of Statistics, North Carolina State College, Raleigh, N. C.; and material for Queries should go to "Queries", Statistical Laboratory, Iowa State College, Ames, Iowa, or to any member of the committee.

As the consideration of articles for publication in *Biometrics* is dependent upon referees, it is necessary for us to have *three* copies of each paper.

We are taking this opportunity to request that in the future all manuscripts be submitted in triplicate.

GERTRUDE M. COX, *Editor*

Membership dues in the American Statistical Association are as follows: regular members $8.00, contributing members $25.00, student members (for a period of 4 years) $4.00, family membership (an additional $2.00 for husband and wife receiving only one copy of the periodicals) $2.00, introductory dues for persons under 30 (for the first year only) $4.00, associate members of the Biometrics Section $4.00, senior members (for persons over 60 having been members of the ASA for 25 years or more) exempt, life membership as voted on October 7, 1948, i.e., 25 to 30 years $200.00, over 30 at Age A $290.00 −3A.

Annual subscriptions are as follows: To the Journal of *The American Statistical Association,* for non-members $8.00; allocation of members dues $5.00; rate to members of associated or affiliated societies $5.00; to *Biometrics* $4.00, *Biometrics* for American Statistical Association Members $3.00; to *The American Statistician* $1.50. Subscription and applications should be sent to the American Statistical Association, 1108 16th Street, N. W., Washington 6, D.C.

Entered as second-class matter, May 25, 1945, at the post office at Washington, D.C., under the Act of March 3, 1879. Biometrics is published four times a year—in March, June, September and December—by the American Statistical Association for its Biometrics Section. Editorial Office: 1603 K Street, N.W., Washington 6, D.C.

ESTIMATION OF BACTERIAL DENSITIES BY MEANS OF THE "MOST PROBABLE NUMBER"*

WILLIAM G. COCHRAN
School of Hygiene and Public Health
The Johns Hopkins University

Presented at a joint session of the Engineering, Laboratory and Statistics Sections of the American Public Health Association, Biometrics Section of the American Statistical Association, and the Biometric Society, New York, October 27, 1949.

INTRODUCTION

THIS PAPER attempts to give a simple account of the concept of the "most probable number" (m.p.n.) of organisms in the dilution method. The concept is quite old, going back to McCrady (4) in 1915, and has been discussed by various writers from time to time, so that little of what I shall present is new. In addition, some advice is given on the planning of dilution series.

The dilution method is a means for estimating, without any direct count, the density of organisms in a liquid. It is used principally for obtaining bacterial densities in water and milk. The method consists in taking samples from the liquid, incubating each sample in a suitable culture medium, and observing whether any growth of the organism has taken place. The estimation of density is based on an ingenious application of the theory of probability to certain assumptions. For a biologist, it is more important to be clear about these assumptions than about the details of the mathematics, which are rather intricate.

ASSUMPTIONS

There are two principal assumptions. In statistical language, the first is that the organisms are distributed *randomly* throughout the liquid. This means that an organism is equally likely to be found in any part of the liquid, and that there is no tendency for pairs or groups of organisms either to cluster together or to repel one another. In practice this implies that the liquid is thoroughly mixed, and if the volume of liquid is not too great some shaking device is usually employed for this purpose.

---

*Paper 254 from the Department of Biostatistics.

The second assumption is that each sample from the liquid, when incubated in the culture medium, is certain to exhibit growth whenever the sample contains one or more organisms. If the culture medium is poor, or if there are factors which inhibit growth, or if the presence of more than one organism is necessary to initiate growth, the m.p.n. gives an underestimate of the true density.

MATHEMATICAL ANALYSIS

In the mathematical analysis we relate the probability that there will be no growth in a sample to the density of organisms in the original liquid. Suppose that the liquid contains V ml., the sample contains v ml., and that there are actually b organisms in the liquid. By the second assumption, there will be no growth if and only if the sample contains no organisms. We will calculate the probability that none of these b organisms is in the sample.

Consider a single organism. By the first assumption, the probability that it lies in the sample is simply the ratio of the volume of the sample to that of the liquid, i.e. v/V. The probability that it is not in the sample is therefore $(1 - v/V)$. Since there is assumed to be no kind of attraction or repulsion between organisms, these two probabilities hold for *any* organism, irrespective of the positions of the other organisms. (Strictly, this requires the additional assumption that the space occupied by an organism is negligible relative to v.) Consequently, by the multiplication theorem in probability, the probability that none of the b organisms is in the sample is

$$p = (1 - v/V)^b.$$

When v/V is small, this is closely approximated by

$$p = e^{-vb/V}$$

where e, about 2.7, is the base of natural logarithms. Finally, since b/V is the density δ of organisms per ml., we have

$$p = e^{-v\delta},$$

where p is the probability that the sample is sterile.

THE CASE OF A SINGLE DILUTION

If n samples, each of volume v, are taken, and if s of these are found to be sterile, the proportion s/n of sterile samples is an estimate of p. Hence we obtain an estimate d of the density δ by the equation

$$\frac{s}{n} = e^{-vd}.$$

This gives

$$d = -\frac{1}{v} \ln\left(\frac{s}{n}\right) = -\frac{2.303}{v} \log\left(\frac{s}{n}\right) \tag{1}$$

where *ln* and log stand for logarithms to base e and to base 10 respectively.

The estimate d is the "most probable number" of organisms per ml. The derivation given here does not reveal why this name has been ascribed to the estimate. In fact, the concept of m.p.n. is scarcely needed for this simple case. We will, however, reexamine the analysis so as to introduce the concept, which becomes useful in the more complex situation where several dilutions are used.

If p is the probability that a sample is sterile, the probability that s out of n samples are sterile is given by the binomial distribution as

$$\frac{n!}{s!(n-s)!} p^s (1-p)^{n-s} \tag{2}$$

Since $p = e^{-v\delta}$, this expression may be written

$$= \frac{n!}{s!(n-s)!} e^{-sv\delta} (1 - e^{-v\delta})^{n-s} \tag{3}$$

If we have obtained s sterile samples out of n, this formula enables us to plot the probability of this event against the true density δ. Such curves always have a single maximum.

A curve of this type suggests a method for estimating δ which is plausible on intuitive grounds. For if we are considering two possible values of δ, it seems reasonable to prefer the one which gives a higher probability to the result that was actually observed. This argument, carried to its conclusion, leads to a choice of the value of δ for which the probability of obtaining the observed result is greatest. It is this value of δ that has been called the "most probable number" of organisms. It can be shown mathematically that this is the value of δ for which $p = s/n$. Consequently the m.p.n. is the same as the estimate previously given.

In practice, more than one dilution is usually needed. The reason is that the precision of the m.p.n. is very poor when the volume v in the sample is such that the samples are likely to be all fertile or all sterile. When all are fertile, the maximum on the probability curve (3) occurs when δ is infinite, so that the estimated density is infinite. When all are

sterile the estimated density is zero, as may also be verified from equation (1). Thus a single dilution is successful only if v happens to be chosen so that some samples are sterile and some are fertile. Such a choice of v can be made only if the density δ is known fairly closely in advance.

If we possess this knowledge, it is best to select v so that the expected number of organisms per sample lies somewhere between 1 and 2. For this choice the expected percentage of sterile samples will lie between 15% and 35%. In default of this knowledge, the practice is to use several dilutions (i.e. several different values of v) in the hope that at least one of them will give some sterile and some fertile samples.

THREE DILUTIONS

The case of three dilutions serves to illustrate the general problem. Let the suffix i indicate the dilution. For the ith dilution the volume of the sample is v_i, and s_i out of n_i samples are found to be sterile. How do we estimate δ from these results?

From equation (1) we can obtain a separate estimate for each dilution: i.e.

$$d_i = -\frac{2.303}{v_i} \log \left(\frac{s_i}{n_i}\right).$$

However, the best way to combine the three estimates d_i into a single value is not obvious. Since, as we have seen, some dilutions give very poor estimates, it is not satisfactory to take the arithmetic mean.

One solution is provided by the m.p.n. concept, which extends easily to this situation. Following the approach used in the previous section, we first write down the probability of obtaining the observed results for any hypothetical value of the true density δ. The observed results are that s_1 samples out of n_1 are sterile at the first dilution, s_2 out of n_2 at the second, and s_3 out of n_3 at the third. The probability that these three events should all happen is the product of three terms, each like expression (3) in the previous section. As before, the graph of this probability against δ shows a single maximum. The value of δ at this maximum is taken as the m.p.n.

The value of the m.p.n. cannot be written down explicitly. The equation which it satisfies is as follows.

$$s_1v_1 + s_2v_2 + s_3v_3 = \frac{(n_1 - s_1)v_1e^{-v_1d}}{1 - e^{-v_1d}} + \frac{(n_2 - s_2)v_2e^{-v_2d}}{1 - e^{-v_2d}} + \frac{(n_3 - s_3)v_3e^{-v_3d}}{1 - e^{-v_3d}}$$

Methods for solving this equation by trial and error have been given by several writers: e.g. Halvorson and Ziegler (3), Barkworth and Irwin (1)

and Finney (2). In laboratories where the numbers of samples n_i and the dilution ratios are standardized, it is convenient to have a table which gives the m.p.n. for all sets of results that are likely to occur. A table is provided in "Standard methods for the examination of water and sewage" (5), for dilution series in which 5 samples are taken at each dilution and there are three 10-fold dilutions. A more extensive table, for dilution ratios of 2, 4, and 10 and any number of levels (except two levels with a 10-fold dilution) is given by Fisher and Yates (6). This is not a table of the m.p.n., but of a different estimate which seems to be just about as precise for series of the size usually conducted in practice. This estimate is derived from the total numbers X and Y of fertile and sterile samples. The quantities $x = X/n$, $y = Y/n$ are entered in the table, from which an estimate of $\log d$ is obtained.

CRITIQUE OF THE M.P.N.

We have seen that the m.p.n. is an estimate of the density of organisms. Considered more generally, it is a *procedure for obtaining estimates,* since the same argument could be applied to other statistical problems. The only justification which I have mentioned for the procedure is that it seems intuitively reasonable. From a reading of the literature I am not certain as to the reasons which led early investigators to select this estimate, though either the intuitive approach or an appeal to a theory of inverse probability may have been responsible.

During the past 25 years the problem of making estimates from data has received much attention from statisticians. Today, most statisticians would, I believe, reject an appeal to intuition or to the theory of inverse probability as a reliable procedure for constructing estimates, since both have been found on occasion to be untrustworthy. They might also object to the name "most probable number," on the grounds that the adjective "probable" in that phrase has a different meaning from the one given to it in the theory of probability. The estimate is "most probable" only in the roundabout sense that it gives the highest probability to the observed results. But they would not reject the m.p.n. procedure itself, which has come to be regarded as a remarkably reliable tool of very wide utility. At the risk of a slight digression it is interesting to indicate the reasons for the reputation which the method has acquired.

The modern approach is to appraise any method of estimation by results. For the m.p.n. this is done, ideally, by conducting a large number of dilution series with given v's and n's, in circumstances where the true density is known. For each series the density is estimated by the m.p.n., so that we accumulate a large number of observations on the amounts by which the m.p.n. is in error. These observations can be

summarized conveniently by plotting the frequency distribution of the m.p.n. about the true density. If this frequency distribution groups very closely about the true density, we know that the estimates are usually good. Such a set of experiments would be difficult and expensive to conduct, but if we assume that the mathematical analysis which has been applied to the dilution method is valid, we can work out the frequency distribution by purely mathematical methods.

As the numbers of samples n_i become large, the frequency distribution of such an estimate (m.p.n. or other) usually tends to assume a certain limiting form—the normal distribution. An important general result has been established about these limiting distributions (7), to the effect that the limiting distribution of the m.p.n. has the smallest standard deviation that can be achieved by any method of estimation. Roughly speaking, this means that the m.p.n. gives on the average at least as precise estimates as any other method used on the same data. There is no point in seeking further for a more precise estimate. The theorem cannot be proved in general when the numbers of samples are small, but experience suggests that the m.p.n. technique is among the best methods of estimation in this case also. Consequently the m.p.n. method is now generally used in a great variety of problems of statistical estimation, though it more frequently goes by the name of the "method of maximum likelihood."

THE PLANNING OF DILUTION SERIES

In preparation for an estimation by the dilution method, three decisions must be made: (i) what range is to be covered: i.e. what are to be the highest and lowest sample volumes; (ii) what dilution factor is to be used; and (iii) how many samples should be taken for each dilution.

Specific decisions must depend on a knowledge of the limits within which the true density is likely to lie and on the precision desired in the estimate. The way in which precision is to be measured needs some comment. Suppose that the true density is thought to lie somewhere between say 2 and 400 organisms per ml. No matter where the true density should happen to be within this range, we want to plan the series so that the estimate will have a specified "precision." This might be taken to mean that the standard error of the estimated density should be say 30 organisms. But this does not seem a reasonable definition of "equal precision," because although an estimate of 360 ± 30 organisms seems satisfactorily precise, an estimate of 5 ± 30 organisms seems very imprecise. Instead, we take "equal precision" to imply that the standard error bears a constant ratio to the true density, in other words that the coefficient of variation of the estimated density is constant. A further

potent reason for adopting this concept is that in a well-designed series the m.p.n. estimates do have approximately the property that the coefficient of variation is independent of the true density. Thus in a sense we are making a virtue of necessity.

The following remarks are intended as a rough guide in the planning of dilution series. They were derived from investigations of the precision of the m.p.n.

HIGHEST AND LOWEST SAMPLE VOLUMES

These are determined by the range of densities with which we expect to have to cope. With a single dilution it was mentioned that for the best results the expected number of organisms in the sample volume v should lie between 1 and 2. It follows that in a series of dilutions the expected number of organisms in the *highest* sample volume v_H should be at least 1, otherwise there is a risk that all samples will be sterile. Similarly the expected number of organisms in the *lowest* sample volume v_L should not exceed 2, to avoid the risk that all samples will be fertile. This line of reasoning would lead to the rule that a dilution series is capable of estimating any density that lies between $1/v_H$ and $2/v_L$.

This rule is satisfactory if a substantial number of samples, say 20 or more, are being taken at each dilution. With very small numbers of samples per dilution, which are typical in certain lines of work, the rule is not quite stringent enough, in that it allows too much risk that all samples may be fertile. Suppose that we have three 10-fold dilutions, with sample volumes 0.01, 0.1 and 1 ml. This series should be able to estimate any true density between 1 and 200 organisms per ml. If, however, the density happens to be 200 per ml., so that the expected number of organisms per sample in the lowest sample volume is 2, then the probability of a sterile sample at this dilution is e^{-2}, or 0.135. The probability of a fertile sample is 0.865. If only four samples are used per dilution, the probability that all four are fertile is $(0.865)^4$, or 0.56. At the two higher concentrations, all samples are practically certain to be fertile. Thus the worker runs about a 50-50 chance that all his samples will be fertile, which usually necessitates repetition of the series. On the other hand, with 20 samples per dilution, the probability that all are fertile is $(0.865)^{20}$, or only about 0.05.

Thus in small experiments it is safer to reduce the upper density value from $2/v_L$ to $1/v_L$. In practice, we use this rule by first guessing two limits δ_L and δ_H between which we are fairly certain that the true density lies. The sample volumes are then chosen to satisfy the rules

$$v_H \geq \frac{1}{\delta_L} ; \qquad v_L \leq \frac{1}{\delta_H} .$$

For example, if we are confident that the density lies between 10 and 750 per ml., the highest sample volume should be at least 1/10, or 0.1 ml. The lowest sample volume should not be more than 1/750 ml. The three 10-fold dilutions 1/10, 1/100 and 1/1000 ml., or the four 5-fold dilutions 1/10, 1/50, 1/250 and 1/1250, would amply cover this range of densities.

THE DILUTION RATIO

As regards the selection of a dilution ratio, there are two relevant results. If the total number of samples in the whole series is kept fixed, the average precision is practically the same for any dilution ratio between 2 and 10. The advantage of a low dilution ratio, which requires more work, is that the precision is more nearly constant throughout the range of densities between $1/v_H$ and $1/v_L$. These points may be illustrated by a comparison between the dilution ratios 2 and 10, in series designed to cover the same range of densities and to use the same total number of samples, 72. The details for the two series are as follows.

Dilution ratio	No. of samples per dilution	Volumes of samples (ml.)
2	9	.01, .02, .04, .08, .16, .32, .64, 1.28
10	24	.01, .10, 1.00

The two series should cover a range of densities from $1/v_H$ to $1/v_L$, or from about 1 to 100 organisms per ml. The dilution ratio 2 requires eight dilutions, with 9 samples per dilution, whereas the dilution ratio 10 requires only 3 dilutions and allows 24 samples per dilution.

In Figure 1 the standard error of the m.p.n., expressed as a percent of the true density, is plotted against the true density (on a log scale). With both dilution ratios the standard error per cent is fairly constant for any true density between 1 and 100 organisms per ml. Outside these limits the standard error begins to rise steeply, except that with the 10-fold series, which has 24 samples per dilution, the rise is postponed until $\delta = 200$, for reasons given in the previous section. Inside the limits the standard error shows a periodic fluctuation which is noticeable with the 10-fold dilution but negligible for the 2-fold. With a 5-fold dilution (not shown), this periodic effect would be just perceptible. It is present with the 10-fold series because practically all the information is contributed by a single dilution. When the true density is about 1.5 or 15

or 150, so that one of the dilutions has about 1.5 organisms per sample, there is a trough, with peaks in the intervening densities where no sample has a density close to this value. With the 2-fold series, several dilutions contribute information and the periodic effect is smoothed out. On the whole, the 2-fold dilution gives a slightly lower standard error over the range from 1 to 100 organisms per ml., the difference being about 7 per cent. For these reasons a low dilution ratio is preferable if the extra work involved can be accomplished easily.

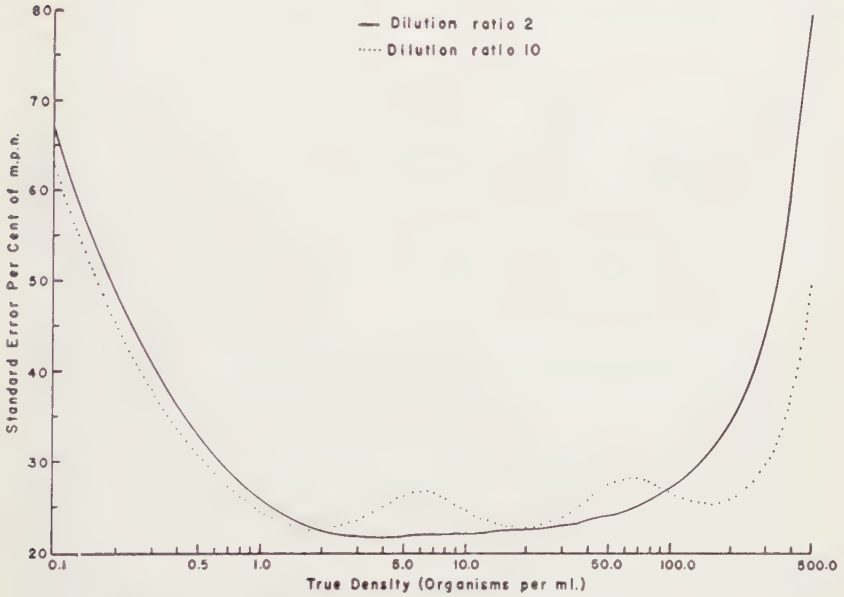


FIGURE I. COMPARISON OF DILUTION RATIOS 2 AND 10

The curves in Figure 1 were calculated by assuming that the formula which holds for the standard error in the limiting distribution, appropriate for very large samples, could be applied to this example in which the total number of samples is 72. Some unpublished work by Dr. I. J. Bross on the distribution of the m.p.n. in small samples indicates that the standard errors are higher than those obtained in this way from the limiting distribution. Further, the periodicity with the 10-fold dilution does not follow the course predicted for it. However, the two principal conclusions from Figure 1 still appear to hold in small samples, namely that the standard error is more stable with a low dilution ratio, and also tends to be slightly lower.*

*This work was carried out under contract with the Office of Naval Research.

STANDARD ERROR OF THE M.P.N.

In many types of investigation there may be only a few samples for each dilution. In this event the distribution of the estimated density d is very skew, and to attach a standard error to d is misleading. The distribution of log d is more nearly symmetrical, and it is recommended that tests of significance and the construction of confidence limits be performed from log d rather than from d. If there are n samples *per dilution* (assumed the same in all dilutions), the standard error of $\log_{10} d$ may be taken as

$$0.55 \sqrt{\frac{\log_{10} a}{n}}$$

where a is the dilution ratio. This formula can be used for any density which lies between $1/v_H$ and $1/v_L$, and for any dilution ratio of 5 or less. For a dilution ratio of 10, a more conservative factor of 0.58 is preferable to 0.55, to allow for the contingency that the estimation may have been made at a point where the standard error has one of its peaks. Thus for dilution ratio 10 the formula becomes simply $0.58/\sqrt{n}$. Note that the formula does not explicitly involve the number of dilutions used.

To test the significance of the difference between two estimated densities, made from independent series, we compute

$$\frac{\log d_1 - \log d_2}{0.55 \sqrt{\frac{\log a_1}{n_1} + \frac{\log a_2}{n_2}}}$$

and refer to the normal probability tables.

The construction of confidence limits may be illustrated by assuming that we have three 10-fold dilutions, with 5 samples per dilution. The standard error of log d is $0.58/\sqrt{5}$, or 0.259, so that the 95 per cent confidence limits for log d are (log $d \pm 0.518$). It follows that to get the upper confidence limit for d, we must multiply d by antilog (0.518) or 3.3, and to get the lower confidence limit we must divide d by 3.3.

For the common dilution ratios, 2, 4, 5, and 10, Table I shows the standard error of log d for any number of samples per dilution between 1 and 10. The table also gives the factor by which the estimated density must be multiplied and divided in order to obtain upper and lower 95 per cent confidence limits respectively. In the example presented by Fisher and Yates (6), the number of rope spore organisms per gram of potato flour was estimated to be 760. The dilution ratio was 2 and there were 5 tubes per dilution. From Table I, the factor for $n = 5$, $a = 2$ is 1.86. Hence the upper confidence limit is 760×1.86 or 1414, while the

TABLE I
STANDARD ERROR OF LOG d AND FACTOR FOR CONFIDENCE LIMITS

No. of samples per dil.	S.E.$_{(\log_{10} d)}$ Dilution ratio (a)				Factor for 95% confidence limits Dilution ratio (a)			
n	2	4	5	10	2	4	5	10
1	.301	.427	.460	.580	4.00	7.14	8.32	14.45
2	.213	.302	.325	.410	2.67	4.00	4.47	6.61
3	.174	.246	.265	.335	2.23	3.10	3.39	4.68
4	.150	.214	.230	.290	2.00	2.68	2.88	3.80
5	.135	.191	.206	.259	1.86	2.41	2.58	3.30
6	.123	.174	.188	.237	1.76	2.23	2.38	2.98
7	.114	.161	.174	.219	1.69	2.10	2.23	2.74
8	.107	.151	.163	.205	1.64	2.00	2.12	2.57
9	.100	.142	.153	.193	1.58	1.92	2.02	2.43
10	.095	.135	.145	.183	1.55	1.86	1.95	2.32

lower limit is 760 1.86 or 409. This factor clearly fulfills the same general purpose as would a standard error, if it had been appropriate to attach one to d.

The table makes it evident that the dilution method is of low precision, as is to be expected from a method that does not use direct counts. Large numbers of samples must be taken at each dilution if a really precise result is wanted. Further, the table is likely to overestimate the accuracy of the method, since it is derived on the assumption that the mathematical analysis corresponds exactly to the practical situation. With a large volume of liquid that cannot be mixed, the distribution of organisms may be far from homogeneous. The method will determine the density in that part of the liquid from which the initial sample was taken. This might be very different from the average density over the whole liquid, and this source of error could be more important than the error in the dilution method itself.

SUMMARY OF STEPS IN PLANNING

The decisions to be made involve a choice of the dilution ratio, a, the number of dilutions and the actual sample volume in each dilution, and finally the number of samples n to be used at each dilution. The steps may be set out as follows.

1. Decide on the limits δ_L and δ_H within which the true density appears certain to lie.

2. Calculate the lowest and highest sample volumes by means of the relations

$$v_H = \frac{1}{\delta_L}, \qquad v_L = \frac{1}{\delta_H}.$$

3. Select a dilution ratio. A low ratio is preferable whenever feasible.

4. The number of dilutions and the actual volumes for each dilution may now be chosen so as to satisfy the requirements that the highest sample volume must not be less than v_H and the lowest must not exceed v_L.

5. The precision to be expected for any specified number n of samples per dilution may be appraised from Table I, if the number of samples per dilution is less than 10, or from the formula for $\text{S.E.}_{(\log d)}$. Choose the number of samples in the light of the precision that is desirable and the amount of work that it is practicable to do.

REFERENCES

(1) Barkworth, H. and Irwin, J. O. (1938). Distribution of coliform organisms in milk and the accuracy of the presumptive coliform test. *J. Hyg.*, Cambridge 38, 446–457.

(2) Finney, D. J. (1947). The principles of biological assay. *J. Roy. Stat. Soc., Ser. B.*, 9, 46–91.

(3) Halvorson, H. O., and Ziegler, N. R. (1933). Application of statistics to problems in bacteriology. *J. Bact.* 25, 101–121.

(4) McCrady, M. H. (1915). The numerical interpretation of fermentation-tube results. *J. Infec. Dis.*, 17, 183–212.

(5) American Public Health Association (1941). *Standard Methods for the Examination of Water and Sewage.* 8th ed.

(6) Fisher, R. A. and Yates, F. (1948). *Statistical Tables for Biological, Agricultural and Medical Research.* Edinburgh, Oliver and Boyd, 3rd ed. Table VIII2.

(7) Fisher, R. A. (1921). On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc.* London, A, 222, 309–368.

STATISTICAL ANALYSIS IN SANITARY ENGINEERING LABORATORY STUDIES*

ROLF ELIASSEN, Sc.D.
Professor of Sanitary Engineering
Massachusetts Institute of Technology
Cambridge 39, Mass.

THE APPROACH of the sanitary engineer to statistics is that of one who is seeking an essential tool for the design of laboratory experiments and the interpretation of their results, and for the planning of the organized collection of field data and the analysis of the significance of this data. The principal distinction between statistical operations on laboratory and field data is that of using small numbers of samples in the laboratory, usually with a few controlled variables, as opposed to the use of very large samples in the field, many times with an appreciable number of uncontrolled variables. This paper will discuss the statistical methods which the writer and his colleagues have used in the planning and execution of laboratory studies.

A sanitary engineering research laboratory is concerned with experimental work involving physical, chemical and biological principles as applied to the treatment of water, sewage and industrial wastes, and the contamination of milk, foods, atmospheres and bodies of water. Many of the phenomena under investigation are biochemical in nature. The vagaries of biochemical reactions, together with the heterogeneity of the media employed, frequently lead to experimental results which may follow a general trend but vary with individual observations. Under these conditions it is obvious that the sanitary engineer and his associates in the fields of chemistry and biology have a great need for the application of statistics in the planning of experiments and evaluation of their results.

In addition to a thorough knowledge of their own field of research, sanitary scientists ought to be conversant with the basic concepts of statistics. They should possess the ability to choose the most effective statistical procedure for the analysis of the data at hand, to scrutinize

*Presented before the Engineering, Laboratory and Statistics Sections at 77th Annual Meeting of the American Public Health Association on October 27, 1949.

the fundamental assumptions underlying the particular statistical procedure chosen, and to recognize by testing to determine whether these assumptions are fulfilled by the situation surrounding the data. Frequently they may have to call upon the services of mathematical statisticians to assist in this work. Statistics then becomes a scientific tool for testing hypotheses and estimating the significance of experimental results obtained in the sanitary science research laboratory.

In the planning of experiments in new fields of sanitary engineering research, it is frequently necessary to develop new analytical techniques. Such a situation arose in connection with a study of the effect of chlorinated hydrocarbons on the suppression of hydrogen sulfide production in sewage (1). The problem involved the sweeping out of hydrogen sulfide from an anaerobic culture medium by means of an inert gas and trapping the hydrogen sulfide for quantitative analysis with the complete exclusion of air.

Repeated trials of the technique, together with necessary revisions, were observed by means of statistical tests, particularly standard deviations and their coefficients of variation. Starting with coefficients as high as 30 per cent, controls were effected until the results in Table I were obtained.

TABLE I

SULFIDES PRODUCED IN ANAEROBIC FERMENTATION OF SEWAGE AFTER 42 HOURS AT 37°C.

Run No.	H_2S ppm	Total Sulfides ppm
1	210	239
2	221	215
3	218	240
4	228	227
5	220	230
6	227	226
7	223	227
8	224	228
9	192	208
Mean	218	227
Standard Deviation ($\bar{\sigma}$)	11.1	10.2
Coefficient of variation (%)	5.1	4.5

These statistical results indicate that two-thirds of the time this analytical procedure can be expected to yield quantitative measures

which will be within 5 per cent of the mean value which would be obtained if a number of runs were made. Furthermore, 95 per cent of the time the results would be within 10 per cent of the mean. Following this, the research work could then be carried out with the assurance that experimental errors in analytical techniques had been reduced to a satisfactory level.

A different type of correlation analysis was involved in the development of a rapid method for the determination of dissolved oxygen in sea water. When making pollution studies in harbors and estuaries, it is preferable that all chemical analyses be made on the sample boat. The Winkler test (2) does not lend itself too readily to such a situation. In 1935 Gilcreas (3) introduced a colorimetric method using Amidol for dissolved oxygen determination in ordinary waters. However, application to the analysis of sea water was not feasible because the presence of sodium chloride interfered with the test. In an attempt to adapt the method to sea water Lieber (4) conducted a laboratory investigation as a master's thesis under the author's direction at the College of Engineering of New York University. The results of many tests with Amidol (diamidophenol hydrochloride) were compared on a statistical basis with those using the Winkler Method in waters having a wide range of salinity.

By the method of least squares the data was fitted to a linear equation of the type $y = a + bx$. This could be used in graphical form to convert the results of field observations with Amidol (y) to equivalent Winkler values (x). From the lines of regression for each salinity value the coefficients of correlation were obtained. These values are indicated in Table II

TABLE II

STATISTICAL RESULTS IN COMPARING AMIDOL AND WINKLER METHODS FOR DISSOLVED OXYGEN DETERMINATION IN SEA WATER

Percent Sea Water	Line of Regression $Y = a + bx$		Standard Error of Estimate	Coefficient of Correlation
	a	b	Sy	r-%
0	−0.223	1.103	0.176	99.25
20	−0.338	1.191	0.230	99.65
40	−0.222	1.325	0.262	99.50
60	−0.123	1.341	0.348	98.99
80	−0.505	1.473	0.367	99.35
100	−0.386	1.536	0.258	99.65

Applying the method of least squares, or the simpler method of graphical analysis, it would be desirable to express the parameters a and b as linear functions of percent salinity. These relationships have been evaluated approximately by the graphical method:

$$a = -0.250 - .001 \ (\% \text{ salinity})$$
$$b = 1.100 + 0.0045 \ (\% \text{ salinity})$$

For a complete analysis, these values of a and b would have to be tested for correlation with observed data and the constants in the above equations revised by trial and error to the degree of accuracy required.

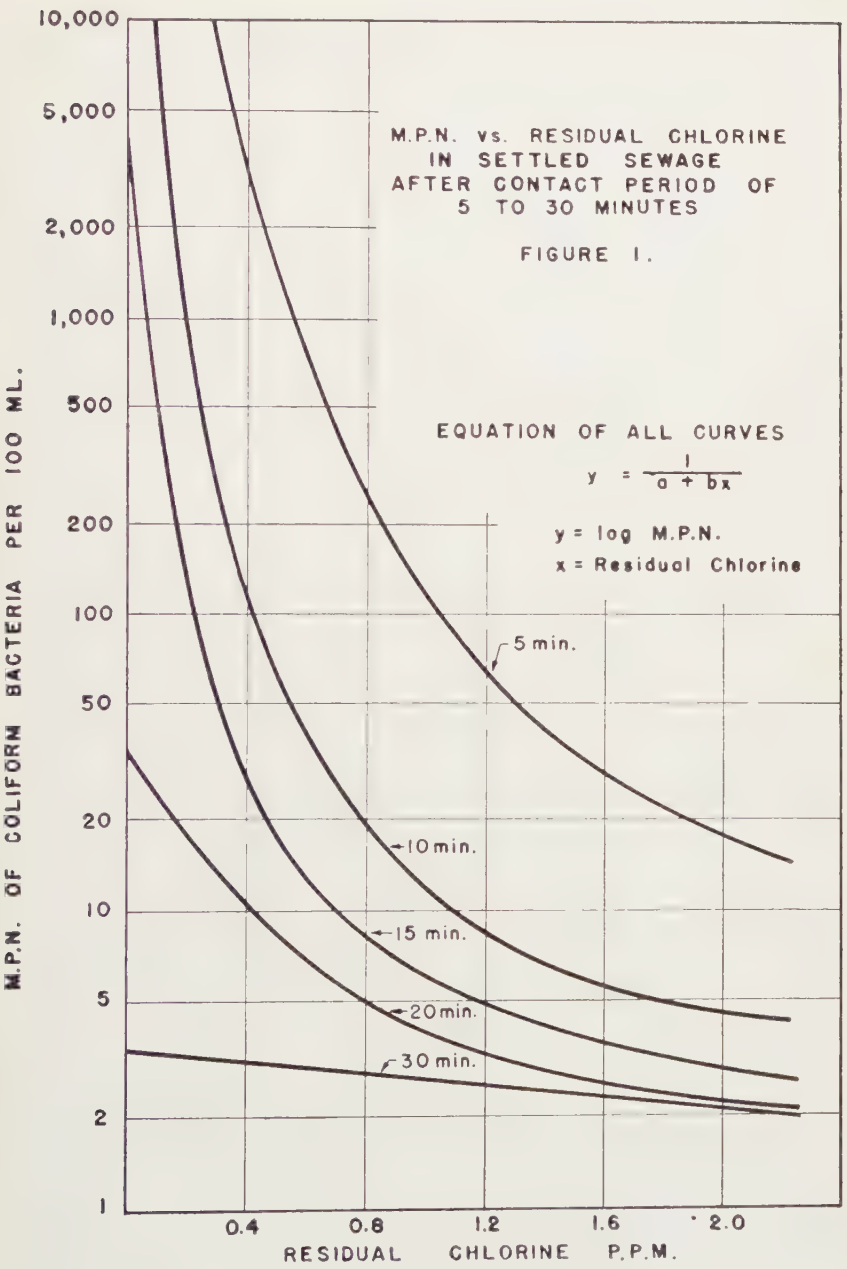
A remarkably high degree of correlation was evident in the fit of the equations determined by least squares and the observed data. For example, the Sy for 60 percent salinity was 0.348. This indicates that the method will yield results within 0.35 ppm about 68 per cent of the time or within 0.70 ppm about 95 per cent of the time. Thus, the Amidol method may be relied upon to yield results with a good degree of reliability, at the same time providing a highly useful tool for pollution studies.

In bacteriological studies the fitting of curves to scatter diagrams is particularly useful for the interpretation of results. On a sponsored research project at New York University, Krieger (5) studied the effect of various chlorine contact periods on the rate of killing of coliform organisms. The observed data indicated that the phenomena could best be interpreted on the basis of a curve of non-linear regression in the form of

$$\log y = \frac{1}{a + bx} .$$

TABLE III
CURVILINEAR REGRESSION CHARACTERISTICS OF M.P.N. VS. CHLORINE RESIDUAL EXPERIMENTS

Contact Time Minutes	Square of Standard Error $(Sy)^2$	Square of Standard Deviation $(\sigma_y)^2$	$\log y = \frac{1}{a + bx}$	
			a	b
5	0.192	0.240	0.17	0.32
10	0.649	0.899	0.21	0.74
15	0.672	1.180	0.28	1.04
20	1.132	1.552	0.65	1.02
30	1.554	1.640	1.91	0.47


M.P.N. vs. RESIDUAL CHLORINE IN SETTLED SEWAGE AFTER CONTACT PERIOD OF 5 TO 30 MINUTES
FIGURE 1.
EQUATION OF ALL CURVES
$y = \frac{1}{a + bx}$
y = log M.P.N.
x = Residual Chlorine
5 min.
10 min.
15 min.
20 min.
30 min.
M.P.N. OF COLIFORM BACTERIA PER 100 ML.
10,000
5,000
2,000
1,000
500
200
100
50
20
10
5
2
1
0.4
0.8
1.2
1.6
2.0
RESIDUAL CHLORINE P.P.M.

The M.P.N. per 100 ml. were represented by y and the chlorine residual (R) after specified contact periods by x in the general equation:

$$\text{M.P.N.} = \exp\left(-\frac{k}{a + bx}\right).$$

By the method of least squares, curves were developed for the various contact periods employed in the experiments. The statistical characteristics of the curves are shown in Table III.

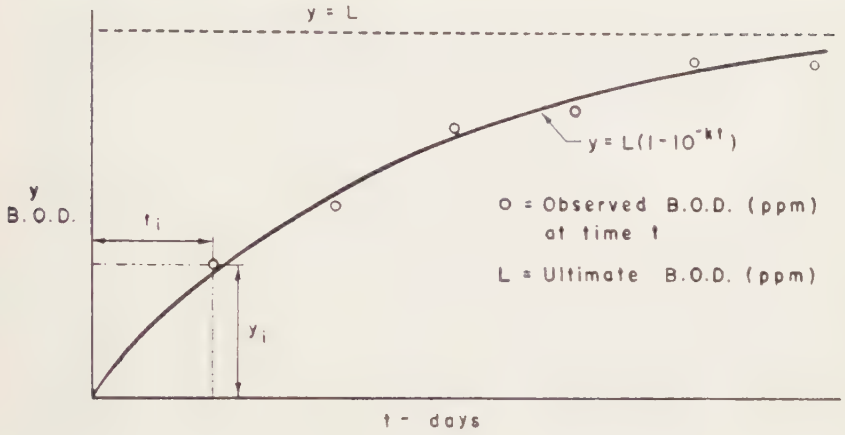
Plotting the values corresponding to the derived equations gave the family of curves shown in Fig. 1. By means of curves of a similar type derived for any particular sewage, state departments of health, designing engineers and plant operators can be guided in the design of chlorine contact chambers and the proper chlorine dosages to achieve the degree of disinfection of the sewage required for the receiving stream.

Laboratory studies of sewage oxidation and stream degradation require extensive use of the biochemical oxygen demand (B.O.D.) test. In spite of the vagaries of biochemical reactions, particularly when used for quantitative measures, the results can be interpreted successfully by conducting a sufficient number of experiments and subjecting the results to statistical analysis.

It has been generally agreed that the first stage biochemical oxidation of sewage is a unimolecular phenomenon proceeding in accordance with the relation $y = L(1 - 10^{-kt})$. L represents the ultimate first stage B.O.D. (ppm), y is the B.O.D. after time t in days, and k is the reaction rate constant. For research workers the determination of the parameters k and L is extremely important. Various statistical procedures may be applied to the observed data. The most accurate method would be by least squares, but for an equation of this type the computations would prove too cumbersome for large amounts of data. The most convenient method, and one which has proven very accurate, is that suggested by H. A. Thomas (6) and known as the Method of Moments.

The statistical hypothesis employed is that the observed data can be fitted to a unimolecular curve by assuming that the zeroth moment and the first moment for the data equal the same moments for the fitted theoretical curve. The type of curve and the equations of moments are shown in Fig. 2. Evaluation of the summations for all of the data may be facilitated by means of graphs prepared by Professor Thomas for specific sequences of days of B.O.D. measurements. By the summation of $\Sigma\, y_i$ and $\Sigma\, y_i t_i$ from the observed data, the parameters k and L may be read from the graphs.

This method has proved to be a useful research tool in an extended



HYPOTHESES

Zeroth Moments $$\sum_{i=0}^{n} t_i^0 y_i = \sum_{i=0}^{n} t_i^0 y$$

First Moments $$\sum_{i=0}^{n} t_i y_i = \sum_{i=0}^{n} t_i y$$

FIGURE 2. METHOD OF MOMENTS FOR FITTING UNIMOLECULAR CURVE TO OBSERVED B.O.D. DATA

series of experiments involving a study of the effect of industrial wastes on the oxidation of sewage and organic matter in streams.

The parameters k and L can be made the basis for further statistical operations in evaluating the significance of results and the difference between means resulting from radioactivity.

In the course of the radioactivity studies at the College of Engineering in New York University, experiments were made to determine whether sewage taken from a specific manhole at the same time on Tuesday and Wednesday could be considered as coming from the same population. Since the ultimate effect of radioactivity would be measured by its effect on the rate factor, k, statistical tests were made on the k values to determine whether there was any significance in the difference between the means.

An hypothesis was set up that the sewage taken on Tuesday and Wednesday would yield no significant difference in k values. As shown in Table IV, t was determined from the computed values of k. Using Student's t distribution, with six degrees of freedom, P was found to be equal to .005. This signifies that one could expect as great, or greater,

difference between the k's for sewage sampled on Tuesday and Wednesday only 5 times in 1000 due to chance alone. Therefore, the hypothesis was rejected.

TABLE IV

t-TEST FOR DIFFERENCE BETWEEN k VALUES OF SEWAGE SAMPLED ON TUESDAY AND WEDNESDAY

k Values		$d = k_W - k_T$	$d - \bar{d}$	$(d - \bar{d})^2$
Tues.	Wed.			
.139	.155	+.016	−.015	.000225
.128	.174	+.046	+.015	.000225
.124	.158	+.034	+.003	.000009
.168	.154	−.014	−.045	.002025
.129	.173	+.044	+.013	.000169
.127	.177	+.050	+.019	.000361
.133	.172	+.039	+.008	.000064

$$\Sigma d = .215 \qquad \Sigma(d - \bar{d})^2 = .003078$$
$$\bar{d} = .031$$

$$\sigma_d = \sqrt{\frac{\Sigma(d - \bar{d})^2}{n - 1}} = \sqrt{\frac{.003078}{6}} = .02265$$

Standard Error of $\bar{d}$, $\qquad \sigma_{\bar{d}} = \frac{\sigma_d}{\sqrt{n}} = \frac{.02265}{\sqrt{7}} = .0085$

$$t = \frac{(\bar{d} - m)}{\sigma_{\bar{d}}} = \frac{.031}{.0085} = 3.6 \qquad \text{Hypothesis: } m = 0$$

Degrees of Freedom = 6

From Student's t Distribution, $\qquad P = .005$

On the basis of these statistical results, the experiments which followed had to be planned to be conducted only on sewage taken at a certain hour on a specific day for each successive run.

Further planning of the radioactivity studies involved the determination of the number of successive experiments to be conducted under similar conditions in order to arrive at mean values of k within the limits

of experimental error. On the basis of experience with the B.O.D. test, it was decided that a standard error of the mean of less than 0.008 (5% of mean k) would be satisfactory. Ten runs were made, with values of k obtained as noted in Table V. Computations indicated that a

TABLE V

NUMBER OF RUNS NECESSARY TO ASSURE MEAN k VALUES WITH A STANDARD ERROR OF LESS THAN .008

k	$k - \bar{k}$	$(k - \bar{k})^2$
.1240	−.0252	.0006
.1275	−.0217	.0005
.1295	−.0197	.0004
.1375	−.0117	.0002
.1390	−.0102	.0001
.1475	−.0017	.0000
.1510	+.0018	.0000
.1680	+.0188	.0004
.1705	+.0213	.0005
.1975	+.0483	.0023
1.4920		.0050

$$\bar{k} = 0.1492 \qquad \Sigma(k - \bar{k})^2 = .005$$

$$\bar{\sigma}_k = \sqrt{\frac{\Sigma(k - \bar{k})^2}{n - 1}} = \sqrt{\frac{.005}{9}} = 2.4 \times 10^{-2}$$

$$\sigma'_{\bar{k}} = \frac{\sigma_k}{\sqrt{N}} = \frac{.024}{\sqrt{10}} = .0076$$

standard error of 0.0076 could be expected with ten separate experiments. Therefore, subsequent experimental work was planned on the basis of ten B.O.D. runs for each concentration of radioisotope in the sewage.

Many other examples might be presented to show the need for and use of statistical methods in sanitary engineering laboratory studies. The author has attempted to show some procedures which have been employed in laboratory work conducted under his supervision. As research workers become more familiar with statistical methods of analysis, they will recognize many places where statistics can be of great

assistance in the planning, execution and interpretation of laboratory work.

Its most valuable place will be found in the interrelation of the oft-times scattered, but nevertheless consistent, data which frequently is obtained in the study of biochemical reactions such as the disinfection of water and sewage, the oxidation of organic matter in sewage treatment processes and in the self-purification of streams, and in anaerobic fermentation processes for sludges resulting from the treatment of sewage and industrial wastes.

REFERENCES

(1) Eliassen, R., Heller, A. N., and Kisch. The Effect of Chlorinated Hydrocarbons on Hydrogen Sulfide Production, *Sewage Works Journal*, 21: 457 (May) 1949.

(2) *Standard Methods for the Examination of Water and Sewage*, 9th Edition A.P.H.A. 1946.

(3) Gilcreas, F. W. A Colorimetric Method for the Determination of Dissolved Oxygen, *Journal of the A.W.W.A.* 27: 1166, 1935.

(4) Lieber, Maxim, and Eliassen, Rolf. The Use of the Amidol Test for Dissolved Oxygen in Sea Water. *Sears Found. Journ. of Marine Research*, 8: 107 (April) 1949.

(5) Krieger, Herman. Master's Thesis, College of Engineering, New York University, May 1948.

(6) Thomas. Lecture Notes in Graduate School of Engineering, Harvard University, May 1946.

(7) Grune, W., Luckens, M., and Eliassen. Report of Research Division, College of Engineering, New York University, Sept. 1949.

THE APPLICATION OF STATISTICAL TECHNIQUES TO SEWAGE TREATMENT PROCESSES*

John W. Fertig
Columbia University
and
Austin N. Heller
The Barrett Division, Allied Chemical & Dye Corporation, New York

Because of the inherent variability in the characteristics of sewage and because of the shortcomings of many of the primary analytical methods used, it is often impossible to determine the specific effectiveness of a given treatment process by examining a single or even a few samples of sewage. It is often necessary to perform many experiments. Because the resultant data may vary considerably, it becomes necessary to apply statistical methods for their evaluation.

There are many phases of sewage treatment processes which can profit from the use of statistical methods. We have chosen to illustrate some of the techniques in terms of the problem of disinfection of raw sewage and sewage effluents.

The results of an extensive series of experiments designed to measure some of the chlorination aspects have been reported by Eliassen, Heller and Krieger (1). It is not our present purpose to review this work, but rather to take extracts from the original experimental data to illustrate some rather fundamental statistical techniques. We shall also draw on hitherto unpublished data referring to the effect of mechanical mixing on M.P.N. of coliform bacteria of raw sewage.

We should like to emphasize that throughout the presentation the stress shall be on statistical methods rather than on these data. We do not purport to believe that the results quoted here will have any great bearing on the operation of sewage treatment plants. We do believe, however, that some of the considerations we raise will have merit in the design of experiments in permitting a more accurate evaluation of treatment processes.

DETERMINATION OF SANITARY QUALITY

Sanitary quality is often measured by the count of coliform organisms per ml., as estimated by the "most probable number" (M.P.N.)

*Presented at the Annual Meeting of the American Public Health Association in New York October 27, 1949.

technique. The M.P.N. of the chlorinated sewage was determined by the partially confirmed test using brilliant green lactose bile broth as the confirmatory medium. For the unchlorinated sewage, direct planting was made into this medium. All M.P.N. determinations were made on the basis of four ten-fold dilutions using three tubes for each dilution.

Although four dilutions were used, the M.P.N.'s in most instances were determined completely (to two significant figures) by the three critical dilutions. Consequently most of the four-dilution codes were translated into M.P.N. by Hoskins' three-dilution tables (2). In general, it is not difficult to select the three critical dilutions. Thus, if the code (for positive tubes) is 3 3 2 0, one selects 3 2 0 to enter the tables. In some codes, it is not so clear which are the three critical dilutions. Thus, given 3 3 1 1, the proper choice is 3 1 1, which gives the same result as the complete four-dilution code. Occasionally the M.P.N. read from the complete four-dilution code differs in the second significant figure from the value appropriate to the three-dilution code. Thus, 3 1 1 0 gives an M.P.N. of 74, whereas 3 1 1 gives 75 and 1 1 0 gives 73.

Where the code 0 0 0 0 occurred, the convention was followed of assigning the lowest value corresponding to one positive tube. Where the code 3 3 3 3 occurred, it was assigned a value corresponding to no positive tubes at the next higher dilution. Other conventions would result in assigning other values to these "indeterminate" codes. However, with a well-chosen system of dilutions, such codes will not arise very often. Furthermore, if there are a large proportion of such indeterminate codes, the data cannot well be analyzed as a series of measurements.

ACCURACY OF M.P.N.'S BASED ON THREE TUBES AT EACH OF THREE TEN-FOLD DILUTIONS

For a given true density of organisms, the M.P.N. obtained from samples may vary considerably through the errors of chance. This variation was studied for the case of three tubes at each of three dilutions. The probability distribution of M.P.N. corresponding to a number of true densities was calculated directly according to the fundamental formula (2). Relatively few of the 64 possible codes occur with any appreciable frequency for any given true density, so that the calculations were not unduly laborious.

The distribution is discrete and very irregular. The terms are not spaced uniformly and do not increase regularly as the true density is approached from either higher or lower sample values. As Halvorson and Ziegler (3) pointed out for the case of 5 or more tubes, the mean

exceeds the true density and the standard deviation increases with the true density.

The distribution of log M.P.N. is more regular, however. While the distribution is discrete and the terms do not increase regularly as the center is approached, it is more symmetrical than the distribution of M.P.N. The mean of the logs is slightly larger than the log of the true density, but much less so than in the case of M.P.N. The standard deviation of log M.P.N. is remarkably constant, varying from 0.29 to 0.33 for the range of true densities tried (0.1 to 5.0). The value of the standard deviation is in fact very close to the value of 0.30 given by Fisher's method of determining the standard error of maximum likelihood estimates (4). This close agreement is surprising since Fisher's method is strictly applicable only for the case where the number of tubes at each dilution is large.

The distribution of M.P.N.'s based on three tubes at each of four dilutions was also investigated for several true densities. It is only slightly different from that described above for three dilutions.

From the above discussion of the distribution it follows that an M.P.N. based on three tubes at each of three (or four) dilutions may differ considerably from the true value. If we may regard the distribution of log M.P.N. as something like a normal curve with three standard deviations as the limit of variation, then an observed M.P.N. may apparently vary from 1/10 to 10 times the true value due to the errors of chance. Two observed M.P.N.'s would have to differ enormously to indicate a significant difference. With a series of experiments, however, one compares the average M.P.N.'s and these have much greater stability than a single M.P.N.

DAILY VARIATION OF OBSERVED M.P.N.'S

In the latter part of 1946, a series of experiments was run on Bronx sewage with chlorine demands ranging from 1.2 to 3.6 p.p.m. The coliform count was determined under various degrees of chlorination and mixing of the sewage and chlorine solution (1). All of the M.P.N. determinations were made in triplicate. In the table below are given the data for nine different days under conditions of no chlorination and no mixing.

It is immediately apparent that the days vary widely from each other, so that the above set of 27 values cannot be considered as homogeneous. On each day, however, the triplicates should exhibit merely random variation. One may compute a measure of this random variation by averaging the nine daily standard deviations, each of which has two

TABLE 1

TRIPLICATE VALUES OF LOG M.P.N. ON NINE DIFFERENT DAYS, UNCHLORINATED SEWAGE, NO MIXING

Day	Log M.P.N. ($=x$)	Total	Average
1	2.97, 3.08, 2.97	9.02	3.01
2	4.04, 4.38, 3.63	12.05	4.02
3	1.56, 2.15, 2.36	6.07	2.02
4	2.63, 3.36, 3.36	9.35	3.12
5	2.97, 3.18, 2.63	8.78	2.93
6	3.36, 3.63, 3.36	10.35	3.45
7	3.63, 3.36, 3.36	10.35	3.45
8	3.36, 3.36, 3.63	10.35	3.45
9	2.63, 3.36, 2.36	8.35	2.78
Total		84.67	3.14

degrees of freedom. This average has eighteen degrees of freedom and may conveniently be computed as

$$\sigma^2 = \left[\Sigma x^2 - \frac{(9.02)^2 + (12.05)^2 + \cdots + (8.35)^2}{3}\right] \div 18 = 0.1014$$

where Σx^2 is the sum of the squares of the individual values (274.8275).

$$\sigma = \sqrt{0.1014} = 0.32$$

This standard deviation agrees remarkably well with the measure of random variation deduced theoretically (0.29-0.33).

Utilizing the triplicates over a number of other situations—100% chlorination, mixing, etc.—we obtain an estimate of random variation of 0.33, based on 108 degrees of freedom.

From the data above we may also obtain a measure of the variation between the daily averages based on eight degrees of freedom. The variation between daily averages (multiplied by 3) is readily obtained as

$$\left[\frac{(9.02)^2 + (12.05)^2 + \cdots + (8.35)^2}{3} - \frac{(84.67)^2}{27}\right] \div 8 = 0.9355$$

A comparison of this variation with a measure of random variation will test whether the days do in fact have significantly different M.P.N.'s on the average. The comparison is made by the variance ratio or F test (5)

$$F = 0.9355/0.1014 = 9.2$$

For 8 degrees of freedom in the numerator and 18 in the denominator, an F of 3.7 is exceeded only 1% of the time when days are the same. Consequently the observed day to day variation appears to be highly significant. Of course, it is well known that the sewage characteristics throughout the course of a day are just as variable or more so than those from day to day.

Because of the significant variation from one type of sewage to another arising from sampling different days, times or places, it is evidently necessary that any comparison of treatments be so arranged that each treatment is represented with each type of sewage sample.

COMPARISON OF TWO DIFFERENT TREATMENTS

On 10 different days during the early part of 1947, a series of experiments was run on Bronx sewage in order to compare the bactericidal effect of various methods of mixing the chlorine solution and the sewage. For our purpose we shall take the data comparing a rapid mix method (initial rapid mixing of 15 sec.) and a no mix method at 100% chlorination. The samples for determining the M.P.N. were taken after a contact period of 10 minutes with the chlorine solution.

TABLE 2

COMPARISON OF RAPID MIXING WITH NO MIXING OF SEWAGE CHLORINATED 100%

Day	Rapid Mix	No Mix	No Mix − Rapid Mix (=Δ)
1	3.04	3.04	0.00
2	2.38	3.38*	+1.00
3	2.38	3.38	+1.00
4	1.97	2.38	+0.41
5	1.36	3.04	+1.68
6	0.97	1.36	+0.39
7	1.97	2.18	+0.21
8	1.63	2.66	+1.03
9	0.15	2.66	+2.51
10	1.63	3.04	+1.41
Total	17.48	27.12	+9.64
Average	1.75	2.71	+0.96 (= Av. Δ)

*M.P.N. of 2400 arbitrarily assigned to code 3 3 3 3.

In this particular case the superiority of one method over the other is evident immediately since for the 9 days where a difference is present, the "no mix" has the higher value. The probability that 9 differences should have the same sign in the absence of a real factor is 1/256.

Although most of the differences are quite large, only a few of them would be statistically significant per se. The difference between two log M.P.N. values is subject to random variation measured by a standard deviation of

$$\sigma = 0.33\sqrt{2} = 0.47$$

Only three differences are as large as three times this standard deviation.

When testing the significance of the difference between two averages, it is customary to compare the difference with the standard error of the difference. This latter quantity is usually evaluated from the standard errors of the two averages. This is not the proper procedure in this case, however. Since the days differ significantly from each other and each method is represented on each day, one must work directly with the daily differences in the two methods.

The standard deviation of the 10 differences is computed as

$$\sigma = \sqrt{\frac{\Sigma\Delta^2}{10} - (\text{Av. } \Delta)^2} = \sqrt{1.4536 - .9216} = \sqrt{.5320} = 0.73$$

The standard error of the average difference is then computed as

$$\sigma_{\text{Av.}} = \sigma/\sqrt{10} = 0.73/\sqrt{10} = 0.23$$

The average difference of 0.96 between the two methods is thus very significant since it is about 4 standard errors distant from zero. Such an occurrence is very improbable according to the normal curve, if zero is indeed the true difference.*

The standard deviation of the ten daily differences (0.73) arises from the fact that the two methods do not portray the same difference on each day. In fact, the obtained value should theoretically equal $0.33\sqrt{2} = 0.47$. The difference is just barely significant. This indicates that the difference between methods varies from one type of sewage to another. To generalize concerning the superiority of one method over another, it is evidently necessary to make the comparison over a range of different sewages. It is perhaps not necessary to make the experiments on different days, but only with different sewages.

On the basis of a larger series of experiments, Eliassen *et al* (1) report

*The use of the *t*-test instead of the normal curve would rarely alter the decision as to significance.

an even larger difference in average M.P.N. between Rapid Mixing and No Mixing than that reported here.

EFFECT OF PROLONGED RAPID MIXING ON M.P.N. OF UNCHLORINATED SEWAGE

In none of the experiments reported by Eliassen *et al* (1) was there a significant difference between the M.P.N. for rapidly mixed and unmixed non-chlorinated sewage. In an unreported series of experiments on Bronx sewage designed to test the effect of rapid mixing for various periods of time, there is a small but statistically significant increase in the average M.P.N. as the time of mixing increases from 0 sec. to 120 sec. The average log M.P.N. increased from 3.66 to 4.14, corresponding to an increase in M.P.N. from 4600 to 14,000.

The type of data is similar to that presented in the preceding section of this paper, except that more than two averages are being compared. The technique of statistical analysis used was an extension of that discussed above, known as the analysis of variance (5).

EFFECT OF CONTACT PERIOD OF CHLORINE SOLUTION ON BACTERIOLOGICAL KILL

A series of experiments was performed during 1946 to study the relation between bacteriological kill and the period during which the chlorine had been in contact with the sewage. The sewage used in these studies was collected mostly from Ridgewood, N. J., and from the Bronx, the chlorine demand ranging from 1 to 6 p.p.m.

There was a rather high proportion of indeterminate results in this study, i.e., results in which the tubes were either all positive or all negative. This was due to a poor choice of dilutions, usually at the shorter contact times. Consequently, the data cannot be analyzed as a series of measurements as in the previous sections of this paper. For this reason, the comparison of contact periods will be made in terms of the proportion of samples with M.P.N. less than some specified number. A value frequently cited is 3 per ml. and this will be used as a dividing point for our purpose.

Taking the data referring to 100% chlorination, we have 27 samples at each contact time.

. To decide if the proportion of samples with M.P.N. less than 3 varies significantly with contact time, we compute chi-square

$$\chi^2 = \Sigma\left[\frac{(0 - T)^2}{T}\right] = 19.4$$

where 0 refers respectively to the six observed frequencies and T refers to the respective theoretical frequencies deduced on the basis of the over-

TABLE 3
DISTRIBUTION OF 81 SAMPLES OF SEWAGE CHLORINATED 100% ACCORDING TO CONTACT PERIOD AND VALUE OF M.P.N.

	Samples with M.P.N.		
Contact Time (min.)	Less than 3	3 or more	Total
3	7	20	27
10	13	14	27
30	23	4	27
Total	43	38	81

all proportion of samples under 3, i.e., 43/81. Thus, at each contact time, the theoretical frequency under 3 is

$$43/81 \times 27 = 14.3$$

and consequently the theoretical frequency of 3 or more is 12.7.

There are two degrees of freedom for χ^2 since three proportions are being compared. From tables (5) of χ^2 we find that a value of 9.2 is exceeded only 1% of the time when contact periods are the same. Consequently, our value of 19.4 is very significant.

For 120% chlorination we find the same degree of significance, while for 40% and 70% the values are not significant, although the proportions vary in the same way with contact time.

In each case the proportion at 10 min. is slightly higher than that at 3 min., but it cannot be considered significant, even if all the data are considered collectively.

The χ^2 method of treating the data illustrated above appears to be very arbitrary in that some division points must be used. Actually several division points were tried without appreciably altering the results. It should be emphasized, however, that if the data can be treated as measurements, such treatment is preferable to the χ^2 method.

USE OF STATISTICAL METHODS IN STUDY OF OTHER TREATMENT PROCESSES

Some of the methods outlined above should be particularly useful in the evaluation of secondary sewage treatment processes, and modifications thereof. In addition, the effect of industrial wastes on biological processes, both aerobic and anaerobic, can probably be more accurately and readily established if requisite experiments are designed to permit statistical evaluation.

SUMMARY

Various statistical techniques useful in the study of sewage treatment processes have been illustrated in terms of data on M.P.N. of coliform bacteria, particularly as influenced by various phases of chlorination.

(1) Most probable numbers determined from 3 tubes at each of 4 dilutions are subject to large chance fluctuations. Their logarithm exhibits a rather constant variability of approximately 0.30, regardless of the true density of organisms.

(2) The M.P.N. of sewage studied here varies significantly from day to day, as shown by comparing the variation between days with random variation.

(3) Rapid mixing of sewage chlorinated 100% resulted in a much lower M.P.N. on the average than when no mixing is employed, as shown by comparing the average difference with the standard error of the average difference. The difference varies from day to day, suggesting that a comparison of methods should be made over a variety of sewages.

(4) Rapid mixing of unchlorinated sewage resulted in a slight but statistically significant increase in the average M.P.N. as the mixing time increased, as shown by an application of the analysis of variance.

(5) At 100% and 120% chlorination the proportion of samples with M.P.N. less than 3 per ml. increases significantly with the contact period of the chlorine solution, as shown by the χ^2 test. While the results were in the same direction at 40% and 70% chlorination, significance could not be established.

REFERENCES

(1) Eliassen, R., Heller, A. N., and Krieger, H. L. A Statistical Approach to Sewage Chlorination. *Sewage Works Jour.* 20: 1008, (Nov.) 1948.

(2) Hoskins, J. K. Most Probable Numbers for Evaluation of Coli Aerogenes Tests by Fermentation Tube Method. *Public Health Reports* 49: 393, (Mar.) 1934. Reprint No. 1621 (Revised 1940).

(3) Halvorson, H. O. and Ziegler, N. R. Consideration of the Accuracy of Dilution Data Obtained by Using Several Dilutions. *J. Bact.* 26: 559, (Dec.) 1933.

(4) Fisher, R. A. On the Mathematical Foundations of Theoretical Statistics. *Phil. Trans. Series A*, 222: 309, 1922.

(5) Snedecor, G. W. *Statistical Methods*, Fourth Edition, 1946. The Collegiate Press, Inc., Ames, Iowa.

FIDUCIAL INTERVALS FOR VARIANCE COMPONENTS

Irwin Bross
Department of Biostatistics
*The Johns Hopkins University**

NATURE OF THE PROBLEM

Research workers frequently have occasion to estimate variance components, especially in genetic and sampling problems. For example, a 1940 Iowa AAA corn acreage study where two sections were selected from each of 1617 townships has the following analysis of variance of the corn acreage per section:

Source of Variation	d.f.	Mean Square	Expected Mean Square
Between Townships	1616	6511.9	$\sigma^2 + 2\sigma_b^2$
Within Townships	1617	1954.3	σ^2

An unbiased estimate of σ_b^2 is

$$\hat{\sigma}_b^2 = \frac{6511.9 - 1954.3}{2} = 2278.8.$$

A symbolic representation of the above table would be

Source of Variation	d.f.	Mean Square	Expected Mean Square
1	n_1	v_1	$\sigma_1^2 = \sigma_2^2 + a\sigma_b^2$
2	n_2	v_2	σ_2^2

and the corresponding estimate

$$\hat{\sigma}_b^2 = \frac{v_1 - v_2}{a} .$$

Now the cautious research worker would like to have fiducial or

*This work was done at North Carolina State College, Raleigh, North Carolina in partial satisfaction of the requirements for the degree of Master of Science in Experimental Statistics.

confidence limits to place on this estimate. The standard error of $\hat{\sigma}_b^2$ may be estimated by

$$\frac{1}{a}\sqrt{\frac{2v_1^2}{n_1}+\frac{2v_2^2}{n_2}}$$

and when n_1 and n_2 are large, the distribution of $\hat{\sigma}_b^2$ tends to be normal so that fiducial limits can be set in the usual way (1). When n_1 and n_2 are small, however, the distribution of the estimate departs considerably from normal, so it would be desirable to have a more refined procedure for small n_1 and n_2.

Satterthwaite (2) has examined the distribution of $a_1v_1 + a_2v_2$ and suggests that it be approximated by a type III curve with effective degrees of freedom

$$n_e = \frac{(a_1v_1 + a_2v_2)^2}{\frac{a_1^2v_1^2}{n_1}+\frac{a_2^2v_2^2}{n_2}}.$$

He warns that if, as in the case of $\hat{\sigma}_b^2$, the a's are negative, then caution must be exercised in using the approximation. This is evident since negative estimates of σ_b^2 can occur which are incompatible with a type III curve.

This paper will present an approximate fiducial interval (which appears to be useful for n_2 as low as six) which is based on R. A. Fisher's approach to the problem.

APPROXIMATE FIDUCIAL LIMITS

R. A. Fisher (3) approaches the problem from a different viewpoint. Let

$$u = v_1 - v_2 \qquad \Upsilon = \sigma_1^2 - \sigma_2^2$$
$$F = \frac{v_1}{v_2} \qquad \Phi = \frac{\sigma_1^2}{\sigma_2^2}$$

then

$$v_1 = \frac{uF}{F-1} \qquad v_2 = \frac{u}{F-1}.$$

Now let x_1 and x_2 be variables distributed as chi-square with n_1 and n_2 d.f. respectively. Then:

$$v_1 = \frac{x_1\sigma_1^2}{n_1} \qquad v_2 = \frac{x_2\sigma_2^2}{n_2} \qquad \Upsilon = \frac{v_1n_1}{x_1} - \frac{v_2n_2}{x_2}$$

and if we take

$$L = \frac{\Upsilon}{u} = \frac{\dfrac{v_1 n_1}{x_1} - \dfrac{v_2 n_2}{x_2}}{(v_1 - v_2)} = \frac{\dfrac{F n_1}{x_1} - \dfrac{n_2}{x_2}}{F - 1}$$

then

$$P(L \leq L_0) = P\left\{\frac{1}{F-1}\left(\frac{F n_1}{x_1} - \frac{n_2}{x_2}\right) \leq L_0\right\}$$

may be found by integrating out x_1 and x_2 with the region of integration determined by the equation in the parenthesis. If we let

$$z = \frac{F n_1}{(F-1)L_0 + \dfrac{n_2}{x_2}}$$

the region of integration is:

$$0 \leq x_2 \leq \infty$$

$$z \leq x_1 \leq \infty$$

so that

$$P(L \leq L_0) = \int_0^\infty f(x_2)\, dx_2 \int_z^\infty f(x_1)\, dx_1 \tag{1.01}$$

which may be regarded as the basic equation.

Now if we can find an L_0 such that $P(L \leq L_0) = \alpha$ we may obtain our fiducial limits immediately since:

$$P(L \leq L_0) = P\left(\frac{\Upsilon}{u} \leq L_0\right) = P(\Upsilon \leq L_0 u) = P\{a\sigma_b^2 \leq L_0(v_1 - v_2)\}$$

$$= P(\sigma_b^2 \leq L_0 \widehat{\sigma_b^2}).$$

Direct integration is not very easy, while tabulation requires four entries, (n_1, n_2, F, α), and involves interpolation, so that it seems more practical to look for an approximate solution.

The basic equation (1.01) may be solved directly in three cases. If n_2 becomes infinite, then

$$\frac{n_2}{x_2} \to 1 \qquad \text{and} \qquad \lim_{n_2 \to \infty} z = \frac{F n_1}{(F-1)L_0 + 1}$$

so that the second integral does not involve x_2 and

$$P(L \leq L_0) = \left(\int_0^\infty f(x_2)\, dx_2\right)\left(\int_z^\infty f(x_1)\, dx_1\right) = \alpha.$$

Hence, if χ^2_α is the value in the chi-square table, $(n_1\ ,\ \alpha)$, then

$$z = \chi^2_\alpha$$

$$L_0 = \frac{\left\{\frac{Fn_1}{\chi^2_\alpha} - 1\right\}}{F - 1}\ .$$

This may also be written (in terms of F-table entries) as

$$L_0 = \frac{\frac{F}{F'_\alpha} - 1}{F - 1}$$

where F'_α is the entry in the F-table for d.f. $(n_1\ ,\ \infty)$.

A second special case is for $F \to \infty$ where

$$\lim_{F\to\infty} z = \frac{n_1}{L_0}$$

and hence $L_0 = 1\ F'_\alpha$. Finally if $F = F_\alpha$, then $L_0 = 0$ is the solution of (1.01). For if $L_0 = 0$,

$$P(L \leq 0) = P\left\{\left(\frac{F_\alpha n_1}{x_1} - \frac{n_2}{x_2}\right)\frac{1}{F-1} \leq 0\right\} = P\left\{\frac{F_\alpha n_1}{x_1} \leq \frac{n_2}{x_2}\right\}$$

$$= P\left(\frac{n_2 x_1}{n_1 x_2} \geq F_\alpha\right).$$

But $(n_2x_1)\ (n_1x_2)$ is distributed as tabular F, F_t , and hence

$$P(L \leq 0) = P(F_t \geq F_\alpha) = \alpha.$$

Now consider the function

$$L = \frac{\frac{F}{F_\alpha} - 1}{F\frac{F'_\alpha}{F_\alpha} - 1} \qquad F \geq F_\alpha$$

$$L = 0 \qquad F < F_\alpha.$$

Then

$$\lim_{F\to\infty} L = \frac{1}{F'_\alpha}$$

$$\lim_{F\to F_\alpha} L = 0$$

$$\lim_{n_2 \to \infty} L = \frac{\frac{F}{F'_\alpha} - 1}{F \frac{F'_\alpha}{F'_\alpha} - 1} = \frac{\frac{F}{F'_\alpha} - 1}{F - 1}$$

so that L is exact for the limiting cases.

Other functions beside L have this property and have been investigated. However L gave the closest agreement with quadrature solutions for the functions investigated.

Some typical values are given in Table I which may serve to indicate the degree of approximation for the case $n_2 = 12$ and $\alpha = .05$.

TABLE I

n_1 \ $F/F\alpha$	1.5	2.0	3.0	5.0
1	5.02	5.13	5.06	5.05
3	5.17	5.31	5.23	5.10
5	5.21	5.21	5.19	5.14

This means, for example, that if $n_1 = 3$ and $F/F_\alpha = 2$, the research worker would actually be working at the 5.31% level instead of the 5% level if he used the proposed approximation. Discrepancies of this order would not ordinarily lead the research worker very far astray.

APPLICATION OF THE APPROXIMATE FIDUCIAL LIMITS

The approximate fiducial interval can be obtained by multiplying $\hat{\sigma}_b^2$ by $\underline{L}$ to obtain the lower limit, and by $\overline{L}$ to obtain the upper limit (central interval). Suppose, for example, that the 5% lower limit is desired. Three quantities are required:

F The ratio v_1/v_2 obtained from the data.
$F_{.05}$ The entry in a 5% F-table for d.f. n_1 and n_2.
$F'_{.05}$ The corresponding entry for d.f. n_1 and ∞.

If $F \leq F_{.05}$, the lower limit will be taken as zero, since σ_b^2 is a positive quantity. If $F > F_{.05}$, then

$$\underline{L} = \frac{\frac{F}{F_{.05}} - 1}{F'_{.05} \frac{F}{F_{.05}} - 1}.$$

The corresponding upper limit

$$\bar{L} = \frac{\dfrac{F}{F_{.95}} - 1}{F'_{.95}\dfrac{F}{F_{.95}} - 1}$$

may be found by using the fact that

$$F_{.95}(n_1 , n_2) = \frac{1}{F_{.05}(n_2 , n_1)}$$

i.e., the 95% value of F may be found by entering the 5% table with the degrees of freedom interchanged and taking the reciprocal of that quantity. Similarly

$$F'_{.95} = \frac{1}{F_{.05}(\infty , n_1)} .$$

The interval between the two points $\underline{L}\hat{\sigma}_b^2$ and $\bar{L}\hat{\sigma}_b^2$ constitutes an approximate 90% fiducial interval.

NUMERICAL EXAMPLE

For purposes of comparison, the numerical example chosen is one where current methods might be expected to apply. A portion of an analysis of variance table (*C*I21 × *NC*7) from H. F. Robinson's data on hybrid corn (5) will be used.

Source of variation	d.f.	M.S.	*E*(M.S.)
Females in males in blocks	184	.0087	$\sigma_2^2 + 2\sigma_b^2$
Pooled error	234	.0033	σ_2^2

The variance component, σ_b^2 , is the additional variance among paternal half sibs due to female differences. Evidently

$$\hat{\sigma}_b^2 = \frac{.0087 - .0033}{2} = .0027$$

$$F = \frac{.0087}{.0033} = 2.64$$

$$F_{.05} = 1.26 \qquad F'_{.05} = 1.18 \qquad \text{(interpolated)}.$$

Whence

$$\underline{L} = \frac{\dfrac{2.64}{1.26} - 1}{1.18\,\dfrac{2.64}{1.26} - 1} = \frac{1.095}{1.47} = .744$$

and $\underline{L}\hat{\sigma}_b^2 = .00202$. Similarly,

$$F_{.95} = \frac{1}{F_{.05}(234,184)} = \frac{1}{1.27}$$

$$F'_{.95} = \frac{1}{F_{.05}(\infty,\ 184)} = \frac{1}{1.20}$$

$$\overline{L} = \frac{2.64(1.27) - 1}{2.64\,\dfrac{1.27}{1.20} - 1} = 1.31$$

so that

$$\overline{L}\hat{\sigma}_b^2 = .00354.$$

A comparison of the corresponding approximate intervals gives:

Interval Used	Lower 5% limit	Upper 5% limit
Fiducial	.0020	.0035
Normal	.0019	.0035
Chi-square (Satterthwaite)	.0021	.0037

so that, as would be expected due to the large number of degrees of freedom, the three methods lead to similar results.

In small samples the methods do not always lead to similar results as the following data on Plankton hauls (6) indicates. The analysis of variance of vertical hauls (Table III)

Source of Variation	d.f.	M.S.	E(M.S.)
Haul	6	.1011	$\sigma_{GH}^2 + 6\sigma_H^2$
Group × Haul	35	.0208	σ_{GH}^2

leads to an estimate of σ_H^2 of .0134. The interval estimates are given below. The Confidence method will be discussed briefly in the next section.

Interval Used	Lower 5% limit	Upper 5% limit
Fiducial	.00420	.0580
Confidence	.00347	.0604
Chi Square	.00556	.0812
Normal	.00000	.0294

The lower limit of the normal approximation is inconsistent with the F-test which is significant at the 5% point.

The analysis of variance of Haul 327

Source of Variation	d.f.	M.S.	E(M.S.)
Haul	9	.1926	$\sigma_{GH}^2 + 6\sigma_H^2$
Group × Haul	45	.0970	σ_{GH}^2

gives .0159 as an estimate of σ_H^2. The intervals are

Interval Used	Lower 5% limit	Upper 5% limit
Fiducial	.00000	.0690
Confidence	.00000	.0740
Chi Square	.00532	.3090
Normal	.00000	.0414

In this case F is not quite significant at the 5% level so the Chi Square result is spurious. A non-zero lower limit will occur with the Fiducial method if and only if the F-test is significant so that contradictions of this nature are impossible. Moreover, as has been indicated earlier, it may be used even when the degrees of freedom are small.

AN APPROXIMATE CONFIDENCE INTERVAL

A rough confidence interval may readily be obtained by using the fact that when $\sigma_b^2 \neq 0$, F is distributed as ordinary F times σ_1^2/σ_2^2. Hence

$$P\left\{F > F_\alpha \frac{\sigma_1^2}{\sigma_2^2}\right\} = \alpha.$$

By manipulating the quantities within the parenthesis, it follows at once that an exact lower confidence limit is

$$\left(\frac{F}{F_\alpha} - 1\right)\frac{\sigma_2^2}{a}.$$

By substituting a sample value for the parameter, a rough confidence interval may be obtained,

$$\left(\frac{\frac{F}{F_\alpha} - 1}{F - 1}\right)\hat{\sigma}_b^2.$$

This result greatly resembles the fiducial approximation. For the first numerical example previously considered, the confidence interval approximation leads to the interval .0018—.0039. It would appear from considerations of expected values (4) that this rough confidence interval tends to be over-conservative. It is interesting to note that, despite theoretical divergences, confidence and fiducial interval methods tend to lead to approximately the same results here.

REFERENCES

(1) Wishart, J., and Clapham, A. "A Study in the Sampling Technique: The Effect of Artificial Fertilizer of the Yield of Potatoes." *Journal of Agricultural Science*, Vol. 19, 616, 1929.

(2) Satterthwaite, F. E. "An Approximate Distribution of Estimates of Variance." *Biometrics Bulletin*, Vol. 6, 110, 1946.

(3) Fisher, R. A. "The Fiducial Argument in Statistical Inference." *Annals of Eugenics*, Vol. 6, 391, 1935.

(4) Bross, I. "Fiducial and Confidence Limits on Components of Variance." Unpublished thesis. 1948.

(5) Robinson, H. F. "The Measurement and Characterization of Genotypic Variance in Segregating Single Cross Populations of Corn (Zea Mays)." Unpublished thesis. 1948.

(6) Winsor, C. P. and Clarke, G. "A Statistical Study of the Variation in the Catch of Plankton Nets." Sear Foundation: *Journal of Marine Research*, Vol. 3, 1940.

DETERMINING SCALES AND THE USE OF TRANSFORMATIONS IN STUDIES ON WEIGHT PER LOCULE OF TOMATO FRUIT

LeRoy Powers

Principal Geneticist, Bureau of Plant Industry, Soils, and Agricultural Engineering, Agricultural Research Administration, United States Department of Agriculture

In biological research the scale on which the variates have been measured may not be in harmony with the nature of the action of the biological processes bringing about the expression of the character under study, as was pointed out by Wright in 1926 and Fisher, Immer, and Tedin in 1932. In such an event the data will not follow the normal probability integral and, if statistical methods of analyzing the data assuming normal distribution are to be used, some transformation of the data causing it to follow the normal probability integral is essential. Mather 1949 suggests that the use of special scales employing some metric or metrics that would be in harmony with the nature of the forces bringing about the expression of the character might furnish a solution. The use of a conventional scale, and if necessary some transformation, is preferred from the standpoint that the equipment for making the measurements in the case of the conventional scale is more readily available and usually does not require any special training for its use. Then the problem resolves itself into determining the scale that is in harmony with the nature of the processes bringing about the expression of the character undergoing investigation and if necessary employing a transformation that will cause the data to follow the normal probability integral.

This paper is concerned with determining scales and the use of transformations in studies on weight per locule of tomato fruit. The locule of the tomato fruit is the chamber containing the seed, its juices, and the placental tissue. Weight per locule was calculated by dividing average weight per fruit of each plant by average number of locules. Hence, the weight per locule also includes other tissues of the tomato fruit, such as the fleshy part which is composed of the locular walls. The data are extensive, involving a large number of plants of four hybrids and covering a period of three years.

In the present study the genetic design of the experiment included P_1 , B_1 to P_1 , F_1 , F_2 , B_1 to P_2 , and P_2 tomato populations and the field design of the experiment was a randomized complete block with 20 replications. The experiments conducted in 1938 had 24 plants planted per plot, and two plots of each of the segregating populations (B_1 to P_1 , F_2 , and B_1 to P_2) were grown per replication to a single plot of any nonsegretating population (parents and F_1). In 1939 and 1940 only 12 plants were planted per plot, and one plot of each population was grown per replication.

EXPERIMENTAL RESULTS

In analyzing data to determine the appropriate scale or scales all constants must be calculated from the individual plant data, and not from frequency distributions. Likewise, transformations should be applied to the individual plant data. Another precaution involves the use of chi-square for testing goodness of fit between the obtained and theoretical frequency distributions. In case the theoretical frequency distribution is calculated from the mean and standard error of the individual plant data from which the obtained frequency distribution is constructed, then chi-square calculations must be based on the differences between the theoretical and obtained frequency distributions and not upon a common frequency distribution such as is used when the theoretical frequency distribution is calculated from a provisional genetic hypothesis based on the data from the nonsegregating populations.

Scale Primarily Logarithmic

The data from the Danmark (*Lycopersicon esculentum* Mill.) × Red Currant (*L. pimpinellifolium* Mill.) tomato hybrid furnish an example of a scale primarily logarithmic. The first step in determining the scale capable of describing the nature of the action of the processes responsible for variation of a character is to compare the obtained means with those calculated on the basis of an arithmetic progression and those calculated on the basis of a geometric progression.

The obtained means and standard errors, theoretical arithmetic and geometric means, grand total variances, and total number of individuals in the study on weight per locule of tomato fruit are given in table 1. In every instance the theoretical arithmetic means are larger than those obtained, whereas, the theoretical geometric means are not significantly different from those obtained. These results indicate that the effects of the forces responsible for the variability of weight per locule of the fruit in the populations of the Danmark × Red Current hybrid are multiplicative. Then, the appropriate scale is logarithmic.

TABLE 1

OBTAINED MEANS AND STANDARD ERRORS, THEORETICAL ARITHMETIC AND GEOMETRIC MEANS, GRAND TOTAL ARITHMETIC VARIANCES, AND TOTAL NUMBER OF INDIVIDUALS FOR WEIGHT PER LOCULE OF FRUIT; DANMARK × RED CURRANT TOMATO HYBRID GROWN IN 1938*

Population	Mean			Grand total arithmetic variance	Number of individuals
	Obtained	Theoretical			
		Arithmetic	Geometric		
	Grams	Grams	Grams		
Red Currant	0.45 ± 0.017			0.017583	420
B_1 to Red Currant	.97 ± .045	2.92	0.99	.183857	932
F_1, D. × R. C.	2.33 ± .130	5.40	2.16	.759823	475
F_2, D. × R. C.	2.12 ± .105	5.40	2.16	1.332184	932
B_1 to Danmark	4.82 ± .253	7.88	4.73	5.213842	928
Danmark	10.36 ± .581			18.286069	457

*In this table and all following tables which involve the Danmark × Red Currant hybrid, D. signifies Danmark and R. C., Red Currant.

As a consequence of these findings, the individual plant data for weight per locule of the fruit of these hybrid and parental populations were transformed to logarithms. The obtained means and standard errors, theoretical logarithmic means, and grand total logarithmic variances of weight per locule are listed in table 2. After transformation, the theoretical arithmetic formulae are used in calculating the theoretical means of the logarithms. The reason for so doing is that multiplicative effects are additive on the logarithmic scale. As was to be expected the theoretical means are not significantly different from those obtained. Since the two parental means were used in calculating the theoretical means of the other four populations given in tables 1 and 2, these figures furnish only indirect evidence as to the nature of the action of the forces causing environmental variability of the parental plants.

To obtain further information concerning the scale appropriate for describing the nature of the action of the forces causing the variability noted for weight per locule of tomato fruit, the data were classified into frequency distributions. The obtained frequency distributions for weight per locule for the two parental and F_1 populations are shown in table 3. In this table the upper class limits are given for both the arithmetic and logarithmic scales. If the appropriate scale is logarithmic then the frequency distributions of table 3 should follow the normal probability

TABLE 2

OBTAINED LOGARITHMIC MEANS AND STANDARD ERRORS, THEORETICAL LOGARITHMIC MEANS, AND GRAND TOTAL LOGARITHMIC VARIANCES FOR WEIGHT PER LOCULE OF FRUIT; DANMARK × RED CURRANT TOMATO HYBRID GROWN IN 1938

Population	Mean		Grand total logarithmic variance
	Obtained logarithmic	Theoretical logarithmic	
Red Currant	−0.364333 ± 0.018357		0.018692
B_1 to Red Currant	−0.051210 ± 0.014673	−0.029018	0.033374
F_1, D. × R. C.	0.334631 ± 0.026734	0.306296	0.031875
F_2, D. × R. C.	0.272647 ± 0.014645	0.306296	0.045940
B_1 to Danmark	0.635670 ± 0.017059	0.641611	0.043278
Danmark	0.976926 ± 0.026607		0.035156

integral. The theoretical frequency distributions based on the assumption that the data do follow the normal probability integral can be calculated from the obtained means and the standard errors of a single determination, which in turn are calculated from the grand total variances.

The grand total variances for the populations must be used in estimating the standard errors of a single determination to be employed in calculating theoretical frequency distributions, because the obtained frequency distributions include the variability due to replications as well as all other variability within a population. The arithmetic grand total variances are given in table 1 and the logarithmic in table 2. The details of calculating the theoretical frequency distributions will be given later in connection with the same calculations for the segregating populations. Chi-square values were calculated to determine whether the deviations noted between the obtained frequency distributions and the theoretical frequency distributions were greater than expected on the basis of random sampling.

The degrees of freedom, chi-square values, and P values for testing goodness of fit between the obtained frequency distributions and those calculated on the basis of the arithmetic and logarithmic scales are given in table 4. Since all of the values for weight per locule of the fruit for the Red Currant parent fell in the first class of the frequency distribution, there are no values for this parent. All of the chi-square values for the theoretical frequency distributions calculated on the arithmetic scale are too large to be attributable to probable errors of random sampling. The P value of the chi-square calculated from the F_1 obtained and

TABLE 3

OBTAINED FREQUENCY DISTRIBUTIONS OF WEIGHT PER LOCULE OF FRUIT FOR THE TWO PARENTAL AND F_1 POPULATIONS; DANMARK × RED CURRANT TOMATO HYBRID GROWN IN 1938

Population	Frequency distribution																								Total number of individuals
	Upper class limit in grams (number and logarithm)																								
	1.5 0.176091	2.5 0.397940	3.5 0.544068	4.5 0.653212	5.5 0.740363	6.5 0.812913	7.5 0.875061	8.5 0.929419	9.5 0.977724	10.5 1.021189	11.5 1.060698	12.5 1.096910	13.5 1.130334	14.5 1.161368	15.5 1.190332	16.5 1.217484	17.5 1.243038	18.5 1.267172	19.5 1.290035	20.5 1.311754	21.5 1.332438	22.5 1.352182	23.5 1.371068	24.5 1.389166	
	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.	no
Red Currant	420	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	420
F_1, D. × R. C.	83	225	131	31	3	1	1	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	475
Danmark	..	1	6	21	24	41	47	34	40	35	44	40	32	21	17	15	4	15	6	7	3	..	1	3	457

TABLE 4

DEGREES OF FREEDOM, CHI-SQUARE, AND P VALUES FOR TESTING GOODNESS OF FIT BETWEEN THE OBTAINED FREQUENCY DISTRIBUTIONS AND THOSE CALCULATED ON THE BASES OF THE ARITHMETIC AND LOGARITHMIC SCALES FOR WEIGHT PER LOCULE OF FRUIT FOR THE P_1, F_1 AND P_2 POPULATIONS; DANMARK × RED CURRANT TOMATO HYBRID GROWN IN 1938

Population	Degrees of freedom		Chi-square		P lies between	
	Arithmetic	Logarithmic	Arithmetic	Logarithmic	Arithmetic	Logarithmic
Red Currant	0	0	—	—	—	—
F_1, D. × R. C.	3	4	11.792	14.142	0.01 & —	0.01 & —
Danmark	16	16	48.393	29.546	0.01 & —	0.05 & 0.02

logarithmic frequency distributions is less than 0.01, and P for the corresponding calculations for the Danmark population is less than 0.05 but larger than 0.02. Another population of Danmark was grown in this same experiment, but in connection with the Danmark × Johannisfeuer cross. Chi-square calculated from the obtained frequency distribution and the theoretical logarithmic frequency distribution had a P value lying between 0.50 and 0.30, showing that the proper scale must be logarithmic for the data from this parent. All of the chi-square values listed in table 4 can be calculated from the data given in tables 1, 2, and 3.

If the genetic variability is due to multiplicative effects of the forces

TABLE 5

MEANS, GRAND TOTAL VARIANCES, STANDARD ERRORS OF A SINGLE DETERMINA OF FRUIT; SEGREGATING POPULATIONS OF THE DANMARK × RED CURRANT TRANSFORMED TO LOGARITHMS

Population	Mean	Variance	Standard error	Frequency Distribution				
				Upper limit of class				
				0.176091	0.397940	0.544068	0.653212	0.740363
				%	%	%	%	%
B_1 to Red Currant . . .	−0.051210	0.033374	0.182686	89.3	10.0	0.6	0.1	..
F_2, D. × R. C.	0.272647	0.045940	0.214336	32.6	39.3	17.9	6.4	2.3
B_1 to Danmark	0.635670	0.043278	0.208034	1.4	11.3	20.3	20.2	15.9

causing it, as was indicated by a comparison between the obtained and theoretical means given in tables 1 and 2, and if such is true for the environmental variability also, then the obtained frequency distributions of the segregating populations should not deviate from their corresponding theoretical logarithmic frequency distributions further than can be explained by chance. The means, standard errors of a single determination, and theoretical frequency distributions for weight per locule of fruit for segregating populations of the Danmark × Red Current hybrid are given in table 5. The theoretical frequency distributions given in table 5 were calculated from the means and standard errors of a single determination by use of Pearson's (1930) tables of the normal probability integral, namely, that portion of table II giving the area and ordinate in terms of the abscissa. In using this table of Pearson's the value under the column heading "x" is obtained by subtracting any given class from the mean and dividing the remainder by the standard error. As an example, take the B_1 to Danmark population given in table 5 and the class having the column heading 0.397940. We have (0.634670 − 0.397940) divided by 0.208034, which results in an "x" value of 1.14. Looking this value up under column heading "x" of Pearson's (1930) table II, page 3, we find the value $1/2(1 + a)$ to be 0.873. Since the value of "x" is positive this must be subtracted from 1.0 and multiplied by 100 to give the percent of the population having a logarithm equal to or less than 0.397940. The value obtained is 12.7 percent. Since 1.4 percent of the population is expected to fall in the preceding class, the proportion of the population expected in the 0.397940 class is 12.7 percent minus 1.4 percent, or 11.3 percent as

TABLE 5—*Continued*

TION, AND THEORETICAL FREQUENCY DISTRIBUTIONS FOR WEIGHT PER LOCULE TOMATO HYBRID GROWN IN 1938; INDIVIDUAL PLANT DATA FOR WEIGHT IN GRAMS

Frequency Distribution																
Upper limits of class																Number of individuals
0.812913	0.875061	0.929419	0.977724	1.021189	1.060698	1.096910	1.130334	1.161368	1.190332	1.217484	1.243038	1.267172	1.290035	1.311754	1.332438	
%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	
..	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	932
0.9	0.4	0.1	0.0	0.1	...	...	...	...	...	...	...	...	...	...	...	932
11.1	7.3	4.6	2.8	1.9	1.1	0.8	0.4	0 3	0.2	0.1	0.1	0.1	0.0	0.0	0.1	928

shown in table 5. For a second example, take the class headed 0.977724. We have (0.635670 − 0.977724) divided by 0.208034, which gives an "x" value of −1.64. The value of $1/2(1 + a)$ for the "x" value of 1.64 is 0.949 (Pearson's table II). Since the "x" value is negative this value is not subtracted from 1.0 but is multiplied by 100 to give 94.9 percent, which is the percent of the population having a logarithm less than or equal to 0.977724. Since 92.1 percent of the population is expected to fall in classes having a lower value, 94.9 − 92.1, or 2.8 percent of the population is expected to fall in the 0.977724 class. The values for all of the other theoretical frequency distribution tables, whether on a logarithmic or arithmetic scale, were calculated in a manner identical to the above given examples.

The theoretical frequency distributions given in table 6 were calculated by multiplying the percent expected in any class by the total number of individuals in the population. Thus the theoretical number for the 0.176091 class of the B_1 to Danmark population is 1.4 percent of 928, or 13 individuals. With the exception of the frequency distribution used in calculating the chi-square value of the B_1 to Red Currant population, in which classes 1 and 2 were combined, the classes were grouped so that any given theoretical frequency distribution did not have less than 10 individuals in any class. This grouping was started at the extremes of the frequency distributions as indicated in table 6. In case of the exception noted (B_1 to Red Currant population) the 0.544068 and 0.653212 classes were grouped, even though after so doing the theoretical frequency distribution had only 7 individuals in the last class.

An examination of table 6 reveals that when classes 1 and 2 are not combined the deviations between the obtained and theoretical frequency distributions for the B_1 to Red Currant population are greater than expected, due to the probable errors of random sampling. This is shown by the high chi-square value (10.765) and its corresponding P value which is less than 0.01. The corresponding chi-square value for the B_1 to Danmark population is also somewhat large. In fact, for every population of table 6 the 0.176091 class (first class) had more individuals in the obtained frequency distributions than in the theoretical. The reverse is true for the 0.397940 class (second class). It seems that the data transformed to logarithms are not normally distributed for classes 1 (0.176091 class) and 2 (0.397940 class). The chi-square values with their corresponding P values show that when the first and second classes are combined the deviations between the obtained and theoretical frequency distributions are no greater than expected due to chance fluctuations, as all of the P values lie between 0.30 and 0.20.

TABLE 6

OBTAINED AND THEORETICAL FREQUENCY DISTRIBUTIONS AND CHI-SQUARE FOR TESTING GOODNESS OF FIT FOR WEIGHT PER LOCULE OF FRUIT; SEGREGATING POPULATIONS OF DANMARK × RED CURRANT TOMATO HYBRID GROWN IN 1938; INDIVIDUAL PLANT DATA OF WEIGHT IN GRAMS TRANSFORMED TO LOGARITHMS

Population	Frequency distribution																Chi-square		P lies between	
	Upper limit of class																Classes 1 and 2 not combined	Classes 1 and 2 combined	Classes 1 and 2 not combined	Classes 1 and 2 combined
	0.176091	0.397940	0.544068	0.653212	0.740363	0.812913	0.875061	0.929419	0.977724	1.021189	1.060698	1.096910	1.130334	1.161368	1.190332	1.217484				
	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.	no.				
B_1 to Red Currant																				
Obtained	863	59	8	2	..	..	..	..	..	..	..	..	..	..	..	..	10.765	1.295	0.01 & —	0.30 & 0.20
Theoretical	832	93	6	1	..	..	..	..	..	..	..	..	..	..	..	..				
F_2, D. × R. C.																				
Obtained	327	370	147	51	25	12	..	..	..	..	..	..	..	..	..	..	6.577	5.881	0.30 & 0.20	0.30 & 0.20
Theoretical	304	366	167	60	21	14	..	..	..	..	..	..	..	..	..	..				
B_1 to Danmark																				
Obtained	22	97	187	165	169	120	66	37	27	15	11	5	2	2	2	1	18.795	11.963	0.10 & 0.05	0.30 & 0.20
Theoretical	13	105	188	187	148	103	68	43	26	18	10	7	12	..	..	..				

From the above analysis the following conclusions can be drawn. The data transformed to logarithms are not normally distributed for the first two classes of the frequency distribution, but are normally distributed after the first two classes have been combined. This fact could not be detected by a study of the means alone. Hence, in determining the scale, or scales, capable of describing the nature of the forces bringing about variation of a character, it is desirable to include in the analysis both the means and the frequency distributions. Primarily the data follow the logarithmic scale, and in making a genetic analysis of the data by the partitioning method (Powers, Locke, and Garrett, 1950) the weights per locule of individual plants should be transformed to logarithms and the first two classes of the frequency distributions combined.

Scale primarily arithmetic

The obtained means and standard errors, theoretical arithmetic and geometric means, and total number of individuals for weight per locule of fruit for the Johannisfeuer × Bonny Best tomato hybrid grown in 1939 are given in table 7. None of the arithmetic means differ significantly from their comparable obtained means. However, without exception the theoretical arithmetic means are of smaller magnitude than their corresponding obtained means. Also, the geometric means are considerably smaller than their comparable obtained means, and for most populations significantly so. Clearly, the obtained means do not follow a geometric progression. Likewise, though not as decisive, the evidence indicates that the obtained means do not follow an arithmetic progression either. This does not rule out either the arithmetic or logarithmic scale, as partial dominance of greater weight per locule of the tomato fruit could be responsible for the rather poor agreements noted. It will be remembered that the formulae for calculating both the theoretical arithmetic and geometric means assume no dominance.

In order to obtain further information as to the nature of the action of the biological processes bringing about the expression of the character weight per locule of fruit in the tomato hybrid populations of the cross Johannisfeuer × Bonny Best, an analysis was made of the frequency distributions. The degrees of freedom, and chi-square and P values for testing goodness of fit between the obtained frequency distributions and those calculated on the bases of the arithmetic and logarithmic scales are listed in table 8. With the exception of the B_1 to Johannisfeuer population the fit is good between the theoretical arithmetic frequency distributions and their corresponding obtained frequency distributions. P is considerably less than 0.01 for the chi-square value calculated from

TABLE 7

OBTAINED MEANS AND STANDARD ERRORS, THEORETICAL ARITHMETIC AND GEOMETRIC MEANS, AND TOTAL NUMBER OF INDIVIDUALS FOR WEIGHT PER LOCULE OF FRUIT; JOHANNISFEUER × BONNY BEST TOMATO HYBRID GROWN IN 1939*

Population	Mean			Number of individuals
	Obtained	Theoretical		
		Arithmetic	Geometric	
	Grams	Grams	Grams	
Johannisfeuer	6.20 ± 0.272			224
B_1 to Johannisfeuer	9.96 ± .510	9.44	8.22	224
F_1, J. × B. B.	13.31 ± .469	12.68	10.90	227
F_2, J. × B. B.	13.01 ± .546	12.68	10.90	222
B_1 to Bonny Best	17.02 ± .563	15.92	14.45	221
Bonny Best	19.15 ± .487			213

*In this table and all following tables which involve the Johannisfeuer × Bonny Best, hybrid J. signifies Johannisfeuer and B. B., Bonny Best.

the theoretical arithmetic frequency distribution and obtained frequency distribution of the B_1 to Johannisfeuer population. The fit between the theoretical logarithmic frequency distribution and the obtained frequency distribution for this population is good, as the P value for the chi-square test lies between 0.70 and 0.50. With the exception of the Bonny Best and B_1 to Johannisfeuer populations the fits between the theoretical arithmetic frequency distributions and the comparable obtained frequency distributions are better than the fits between the theoretical logarithmic frequency distributions and their comparable obtained frequency distributions. In case of the Bonny Best population both chi-square values have a P value that lies between 0.20 and 0.10.

The scale is primarily arithmetic, but it should be pointed out that the data are not as discriminatory as is desired, and more work on some populations is necessary to discriminate between scales. However, in making a genetical analysis of these data by the partitioning method the arithmetic scale is satisfactory for all populations except the B_1 to Johannisfeuer. To avoid metrical bias the weights per locule for the individual plants of this population must be transferred to logarithms. Just why the data for the B_1 to Johannisfeuer follow a logarithmic scale is not clear, especially since the F_2 and Johannisfeuer populations give such a good fit to the arithmetic scale. The fact that the obtained frequency distributions are normal when the data are expressed on one or

TABLE 8

DEGREES OF FREEDOM, CHI-SQUARE, AND P VALUES FOR TESTING GOODNESS OF FIT BETWEEN THE OBTAINED FREQUENCY DISTRIBUTIONS AND THOSE CALCULATED ON THE BASES OF THE ARITHMETIC AND LOGARITHMIC SCALES FOR WEIGHT PER LOCULE OF FRUIT; JOHANNISFEUER × BONNY BEST TOMATO HYBRID GROWN IN 1939

Population	Degrees of freedom		Chi-square		P lies between	
	Arithmetic	Logarithmic	Arithmetic	Logarithmic	Arithmetic	Logarithmic
Johannisfeuer	6	7	6.010	32.484	0.50 & 0.30	0.01 & —
B_1 to Johannisfeuer	12	11	44.158	8.767	.01 & —	.70 & 0.50
F_1, J. × B. B.	11	11	8.956	12.679	.70 & .50	.50 & .30
F_2, J. × B. B.	14	14	11.723	30.456	.70 & .50	.01 & —
B_1 to Bonny Best	14	14	13.084	19.540	.70 & .50	.20 & .10
Bonny Best	14	12	19.349	17.245	.20 & .10	.20 & .10

the other, the arithmetic or the logarithmic scales, together with the fact that the obtained means are larger than either the theoretical arithmetic or logarithmic means proves phenotypic dominance of greater weight per locule of fruit in the Johannisfeuer × Bonny Best tomato hybrid, and also is convincing evidence in support of genetic dominance.

Scales arithmetic and logarithmic

The obtained means and standard errors, theoretical arithmetic and geometric means, and total number of individuals for weight per locule of fruit for Johannisfeuer × Red Currant tomato hybrid grown in 1939 are given in table 9. In every instance the arithmetic means are larger than those obtained and the geometric smaller. By employing the obtained mean of the Red Currant parent and the obtained mean of the F_1 population of the Johannisfeuer × Red Currant hybrid, the theoretical geometric mean is 1.09 grams per locule, whereas the obtained mean is 1.04 ± 0.040 grams. Similarly, by employing the mean of the F_1 population of the Johannisfeuer × Red Currant hybrid and the mean of the Johannisfeuer parent, the theoretical arithmetic mean of the B_1 to Johannisfeuer population is 4.45 grams per locule, whereas the obtained mean is 4.48 ± 0.139 grams. From these results one might expect the appropriate scale for the Red Currant and B_1 to Red Currant populations to be logarithmic, and that for the B_1 to Johannisfeuer and Johannisfeuer populations to be arithmetic.

TABLE 9

OBTAINED MEANS AND STANDARD ERRORS, THEORETICAL ARITHMETIC AND GEOMETRIC MEANS, AND TOTAL NUMBER OF INDIVIDUALS FOR WEIGHT PER LOCULE OF FRUIT; JOHANNISFEUER × RED CURRANT TOMATO HYBRID GROWN IN 1939*

Population	Mean			Number of individuals
	Obtained	Theoretical		
		Arithmetic	Geometric	
	Grams	Grams	Grams	
Red Currant	0.44 ± 0.015			229
B_1 to Red Currant	1.04 ± .040	1.88	0.85	233
F_1, J. × R. C.	2.70 ± .058	3.32	1.65	225
F_2, J. × R. C.	2.12 ± .106	3.32	1.65	225
B_1 to Johannisfeuer	4.48 ± .139	4.76	3.20	230
Johannisfeuer	6.20 ± .272			224

*In this table and all following tables which involve the Johannisfeuer × Red Currant hybrid, R. C. signifies Red Currant and J., Johannisfeuer.

The degrees of freedom, chi-square, and P values for testing goodness of fit between the obtained frequency distributions and those calculated on the bases of the arithmetic and logarithmic scales for weight per locule of fruit for the Johannisfeuer × Red Current hybrid grown in 1939 are given in table 10. Since all of the Red Currant plants fell into one class, this population does not provide any information as to the scale appropriate for describing the nature of the action of the processes bringing about the variability of the character weight per locule. For the B_1 to Red Currant population the chi-square values are those expected, if the appropriate scale is arithmetic and not logarithmic, as was indicated by a study of the obtained and theoretical means. The chi-square values also indicate that the proper scale of measurement for the F_1 generation is arithmetic. The same is true of the Johannisfeuer population, but the B_1 to Johannisfeuer population data do not seem to be following the arithmetic scale, as was indicated by the analysis of the means. The chi-square values show that the data are following the logarithmic scale. The F_2 population data do not seem to be following either scale, as was to be expected since the B_1 to Red Currant followed the arithmetic scale and the B_1 to Johannisfeuer the logarithmic scale.

This behavior of the B_1 to Red Currant and B_1 to Johannisfeuer populations, the former following the arithmetic scale and the latter

the logarithmic scale, needs to be considered further, as the reverse was expected from the analysis of the means. These findings raise the the question whether in case of the B_1 to Red Currant population the nature of the action of the processes causing the genetic variability is most adequately described by the logarithmic scale and the nature of the action of the processes causing the environmental variability is most adequately described by the arithmetic scale. If such were the case, the opposite would be true of the B_1 to Johannisfeuer population, that is, the genetic variability would be following the arithmetic scale and the environmental variability the logarithmic scale. Then, since both genetic and environmental variability would be present in these two populations the obtained frequency distribution would be a combination of the two and therefore might be expected not to give a good fit when tested against either the theoretical arithmetic or theoretical logarithmic frequency distributions. However, since the environmental variability makes up the greater proportion of the variability, as regards weight per locule, and the genetic variability a rather small proportion, the fit between the obtained frequency distribution and the theoretical arithmetic frequency distribution might be good in case of the D_1 to Red Currant population and the fit between the obtained and the theoretical logarithmic frequency distribution good in case of the B_1 to Johannisfeuer population. This would mean that the data are not discriminatory as regards the genetic variability. In other words the environmental variability forms such a large proportion of the total variability as to obscure the scale that is followed by the genetic variability.

The obtained means and standard errors, theoretical arithmetic and geometric means, and total number of individuals for weight per locule of fruit for Danmark $\times$ Johannisfeuer tomato hybrid grown in 1938, 1939, and 1940 are listed in table 11. For the 1938 data, with the exception of the B_1 to Danmark population, the obtained means are larger than either the theoretical arithmetic means or the theoretical geometric means. In case of the exception noted, the theoretical arithmetic mean is larger than the obtained mean, but not significantly so. For the 1939 data there are no exceptions, as in every case the obtained means are larger than the theoretical means, whether arithmetic or logarithmic. For the 1940 data, the differences between the theoretical arithmetic means and their respective obtained means for the B_1 to Johannisfeuer and F_2 populations are not significant, but for the F_1 and B_1 to Danmark populations the theoretical arithmetic means are significantly larger than the obtained means. For the same year the logarithmic means are not materially different from the obtained means for the F_1 and B_1 to Danmark populations. However, the logarithmic means are sig-

TABLE 10

DEGREES OF FREEDOM, CHI-SQUARE, AND P VALUES FOR TESTING GOODNESS OF FIT BETWEEN THE OBTAINED FREQUENCY DISTRIBUTIONS AND THOSE CALCULATED ON THE BASES OF THE ARITHMETIC AND LOGARITHMIC SCALES FOR WEIGHT PER LOCULE OF FRUIT; JOHANNISFEUER × RED CURRANT TOMATO HYBRID GROWN IN 1939

Population	Degrees of freedom		Chi-square		P lies between	
	Arithmetic	Logarithmic	Arithmetic	Logarithmic	Arithmetic	Logarithmic
Red Currant	0	0	—	—	—	—
B_1 to Red Currant	1	1	0.193	5.819	0.70 & 0.50	0.02 & 0.01
F_1, J. × R. C.	2	2	1.665	9.529	.50 & .30	.01 & —
F_2, J. × R. C.	3	4	8.692	17.856	.05 & .02	.01 & —
B_1 to Johannisfeuer	5	6	29.092	7.067	.01 & —	.50 & .30
Johannisfeuer	6	7	6.010	32.484	.50 & .30	.01 & —

nificantly smaller than the obtained means for the B_1 to Johannisfeuer and F_2 populations. Then, the analysis of the means indicates that both arithmetic and logarithmic scales are necessary to describe the nature of the action of the genetic processes differentiating the six populations of the Danmark × Johannisfeuer hybrid for the 3 years of the tests.

The degrees of freedom, chi-square, and P values for testing goodness of fit between the obtained frequency distributions and those calculated on the bases of the arithmetic and logarithmic scales for weight per locule of fruit of the Danmark × Johannisfeuer tomato hybrid grown in 1938, 1939, and 1940 are listed in table 12. For the 1938 data, the theoretical arithmetic frequency distribution and the obtained frequency distribution are in agreement for the Johannisfeuer population and the logarithmic frequency distribution is rejected. For the F_1 and Danmark populations the fit is good between the obtained and theoretical logarithmic frequency distributions, and the theoretical arithmetic frequency distributions are rejected. The obtained frequency distributions of the segregating populations (B_1 to Johannisfeuer, F_2, and B_1 to Danmark) do not, in any case, give a good fit to either the theoretical arithmetic or theoretical logarithmic frequency distributions. In all cases the data are discriminatory. The results are those expected on the basis that the processes bringing about the variation of the character, weight per locule of fruit, are most adequately described for some genotypes by the arithmetic scale and for other genotypes by the logarithmic scale. In

TABLE 11

OBTAINED MEANS AND STANDARD ERRORS, THEORETICAL ARITHMETIC AND GEOMETRIC MEANS, AND TOTAL NUMBER OF INDIVIDUALS FOR WEIGHT PER LOCULE OF FRUIT; DANMARK × JOHANNISFEUER TOMATO HYBRID GROWN IN 1938, 1939, AND 1940*

Year and Population	Mean			Number of individuals
	Obtained	Theoretical		
		Arithmetic	Geometric	
	Grams	Grams	Grams	
1938				
Johannisfeuer	4.61 ± 0.446			452
B_1 to Johannisfeuer	6.72 ± .425	5.94	5.58	928
F_1, D. × J.	7.96 ± .419	7.26	6.76	469
F_2, D. × J.	8.35 ± .467	7.26	6.76	932
B_1 to Danmark	8.32 ± .399	8.59	8.19	921
Danmark	9.92 ± .691			456
1939				
Johannisfeuer	6.20 ± .272			224
B_1 to Johannisfeuer	9.45 ± .426	8.28	7.67	230
F_1, D. × J.	11.81 ± .516	10.36	9.48	209
F_2, D. × J.	10.92 ± .382	10.36	9.48	215
B_1 to Danmark	13.74 ± .476	12.44	11.73	231
Danmark	14.51 ± .360			228
1940				
Johannisfeuer	7.34 ± .062			220
B_1 to Johannisfeuer	10.14 ± .146	9.80	9.08	224
F_1, D. × J.	11.42 ± .148	12.27	11.24	224
F_2, D. × J.	12.41 ± .080	12.27	11.24	223
B_1 to Danmark	13.37 ± .253	14.74	13.90	219
Danmark	17.20 ± .259			219

*In this table and all following tables which involve the Danmark × Johannisfeuer hybrid, D. signifies Danmark and J., Johannisfeuer.

such an event the frequency distributions of the nonsegregating generations would be expected to show a good fit to a frequency distribution based on one or the other scale, whereas the obtained frequency distributions of the segregating generations would not be expected to give a good fit when tested to either one or the other scale. As has been shown such was the situation.

For 1939 the obtained frequency distributions of the Johannisfeuer and F_1 populations gave a good fit when tested for the arithmetic scale

TABLE 12

DEGREES OF FREEDOM, CHI-SQUARE, AND *P* VALUES FOR TESTING GOODNESS OF FIT BETWEEN THE OBTAINED FREQUENCY DISTRIBUTIONS AND THOSE CALCULATED ON THE BASES OF THE ARITHMETIC AND LOGARITHMIC SCALES FOR WEIGHT PER LOCULE OF FRUIT; DANMARK × JOHANNISFEUER TOMATO HYBRID GROWN IN 1938, 1939, AND 1940

Year and population	Degrees of freedom		Chi-square		*P* lies between	
	Arithmetic	Logarithmic	Arithmetic	Logarithmic	Arithmetic	Logarithmic
1938						
Johannisfeuer	8	10	11.636	44.326	0.20 & 0.10	0.01 & —
B_1 to Johannisfeuer	13	15	118.466	35.936	.01 & —	.01 & —
F_1, D. × J.	14	13	67.582	6.800	.01 & —	.95 & 0.90
F_2, D. × J.	16	19	136.745	33.310	.01 & —	.02 & .01
B_1 to Danmark	16	17	77.482	47.078	.01 & —	.01 & —
Danmark	16	16	52.572	17.016	.01 & —	.50 & .30
1939						
Johannisfeuer	6	7	6.010	32.484	.50 & .30	.01 & —
B_1 to Johannisfeuer	11	11	8.738	6.244	.70 & .50	.90 & .80
F_1, D. × J.	10	11	4.306	27.528	.95 & .90	.01 & —
F_2, D. × J.	10	10	13.762	8.335	.30 & .20	.70 & .50
B_1 to Danmark	10	12	28.285	17.609	.01 & —	.20 & .10
Danmark	11	10	12.183	6.056	.50 & .30	.90 & .80
1940						
Johannisfeuer	7	7	7.931	1.801	.50 & .30	.98 & .95
B_1 to Johannisfeuer	13	11	55.484	18.524	.01 & —	.10 & .05
F_1, D. × J.	13	11	40.558	9.951	.01 & —	.70 & .50
F_2, D. × J.	13	11	30.241	6.013	.01 & —	.90 & .80
B_1 to Danmark	13	14	22.966	12.876	.05 & .02	.70 & .50
Danmark	15	14	26.493	14.717	.05 & .02	.50 & .30

but not when tested for the logarithmic scale. The data for the Danmark population were not discriminatory as the obtained frequency distribution gave a good fit when tested for either the arithmetic or logarithmic scale. The same situation was found for the B_1 to Johannisfeuer and F_2 populations. The data for the B_1 to Danmark population rejected the arithmetic scale but the fit between the obtained and theoretical frequency distributions was fair when the test was for the logarithmic scale. These results support the contention that for some genotypes the environmental variability is most adequately described by the arith-

metic scale and for other genotypes by the logarithmic scale. Since in some cases the data for 1939 are not discriminatory, a greater number of individuals per population are needed.

For 1940 the obtained frequency distributions in all cases give a good fit to the theoretical logarithmic frequency distributions. However, the data for the Johannisfeuer population are not discriminatory, as the obtained frequency distribution for this population also gives a good fit to the theoretical arithmetic frequency distribution. Then, for 1940 the scale capable of describing the action of the biological processes causing variability in weight per locule of the tomato fruit for the cross Danmark × Johannisfeuer is primarily, if not entirely, logarithmic.

These data from the Danmark × Johannisfeuer cross are conclusive in proving that for some genotypes the appropriate scale for describing the nature of the action of the environmental forces causing variability in weight per locule of fruit is arithmetic and for other genotypes is logarithmic. Also, evidence in support of the contention that the scale is not the same for all years is convincing. In 1938 and 1940 the appropriate scale for plants of the F_1 genotype was logarithmic and for 1939 was arithmetic, and in all years the data for the F_1 were discriminatory.

CONCLUSIONS AND SUMMARY

1. With the exception of the Danmark (*Lycopersicon esculentum*) × Red Currant (*L. pimpinellifolium*) tomato hybrid grown in 1938 the arithmetic and logarithmic scales have been sufficient to describe the nature of the action of the biological processes differentiating weight per locule of tomato fruit.

2. In populations of Danmark × Red Currant hybrid, the environmental variability and genetic variability of all genotypes follow the same scale which is logarithmic.

3. This is not true for the populations of any other hybrid, as the environmental variability of some genotypes was found to be arithmetic and that of others logarithmic.

4. The same was true of the genetic variability for the Johannisfeuer (*L. esculentum*) × Red Currant hybrid, as the mean of the B_1 to Red Currant was geometric and that of the B_1 to Johannisfeuer was arithmetic.

5. When using the partitioning method of analyzing segregating populations, the environmental scale must be employed as the data for any given segregating population is partitioned on the basis of genotypes. By so doing the genetic variability is removed.

6. In those cases where the environmental variability of all genotypes of a hybrid are not following the same scale, the transformations (if any are necessary) must be based on the scales indicated by the environ-

mental variability of each individual genotype.

7. For the populations of any given hybrid the genetic variability may be following one scale and the environmental variability another. This was found to be true for some genotypes of the Johannisfeuer × Red Currant and the Danmark × Johannisfeuer populations.

8. Likewise, the scale is not necessarily the same for all years, as in 1938 and 1940 the environmental variability of the F_1 population of Danmark × Johannisfeuer hybrid followed the logarithmic scale and in 1939 the arithmetic scale.

9. Red Currant, Danmark, and Johannisfeuer were crossed in every possible combination. For the Danmark × Red Currant hybrid the environmental and genetic variabilities were found to follow the logarithmic scale, primarily. Since Johannisfeuer was found to follow the arithmetic scale as regards environmental variability, for the Johannisfeuer × Red Currant and Danmark × Johannisfeuer populations the environmental variability of some genotypes would be expected to follow the arithmetic scale and that of others, the logarithmic scale. Such was found to be the case. Also, the same was true of the genetic variability.

10. Within the range of the genotypes and environments encountered, both the genotype and the environment are factors in determining the scales appropriate for describing the nature of the action of the biological processes causing variability of weight per locule of tomato fruit. This raises the question whether the variability for any given genotype within one replication may be following one scale and within another replication another scale. The data indicate that such was not the case, as all replications for a given genotype seemed to be following the same scale.

11. In some cases the data were not discriminatory, indicating that too few individuals were grown per population. It appears that each population should be composed of at least 400 individuals, and populations of 900 or more individuals are preferred.

LITERATURE CITED

Fisher, R. A., Immer, F. R., and Tedin, O. The genetical interpretation of statistics of the third degree in the study of quantitative inheritance. *Genetics* 17: 107–124. 1932.

Mather, K. *Biometrical genetics*. 158 pp. Dover Publications, Inc. London. 1949.

Pearson, Karl. *Tables for statisticians and biometricians*. Cambridge University Press. London. 1930.

Powers, LeRoy, Locke, L. F., and Garrett, J. C. Partitioning method of genetic analysis applied to quantitative characters of tomato crosses. *U. S. D. A. Tech. Bull.* 998. 1950.

Wright, S. A frequency curve adapted to variation in percentage occurrence. *Jour. Amer. Statist. Assoc.* New Series 154, 21: 162–178. 1926.

QUERIES

77 QUERY: I have been carrying out an analysis of covariance of yield (y) and age (x) in a clonal cacao experiment of split-plot design, the two splits being for 6 clones (C) and for Buddings v Cuttings (T). Table 1 gives the yields and ages of buddings and of cuttings over the six clones and in the six blocks.

As you will observe, clonal yields differ quite considerably and I have therefore partitioned the 5 d.f. for clones (C) orthogonally into:—

TABLE 1

YIELD (KILOGRAMS) AND SUM OF AGES (UNIT: 10 MONTHS)* OF 8 CACAO TREES IN EACH OF 72 PLOTS ARRANGED IN A SPLIT-PLOT EXPERIMENT
t_1 = BUDDINGS, t_2 = CUTTINGS

Block		Clone											
		1		44		55		70		91		95	
		t_1	t_2	t_1	t_2	t_1	t_2	t_1	t_2	t_1	t_2	t_1	t_2
1	Y	18	29	13	12	5	3	10	4	8	7	16	22
	X	78	78	77	60	78	62	78	44	73	62	78	78
2	Y	30	22	9	12	8	8	6	7	3	7	19	25
	X	77	75	78	58	76	74	76	51	49	64	78	75
3	Y	23	25	18	12	19	5	6	16	11	2	16	18
	X	78	78	78	68	78	64	66	67	71	67	78	76
4	Y	17	21	22	26	12	6	9	10	11	4	19	29
	X	78	78	75	74	76	63	76	57	75	68	78	74
5	Y	28	21	16	12	13	8	9	3	12	9	16	18
	X	76	76	78	66	78	65	71	54	71	61	76	71
6	Y	28	34	13	20	13	7	11	17	11	7	15	12
	X	78	75	78	68	78	65	71	68	74	66	78	78
Total		144	152	91	94	70	37	51	57	56	36	101	124
		465	460	465	394	464	393	438	341	413	388	466	452
Total for Clone		296		185		107		108		92		225	
		925		859		857		779		801		918	

*Editor's note: Querist reported yields in grams and ages in months. The rounded numbers contain all the essential information. Readers who do the computations will encounter the results of rounding.

(i) The mean of the 3 good clones (*ICS* 1,44,95) *v* the mean of the 3 poor clones (*ICS* 55,70,91) ($C' - 1$ d.f.)
(ii) Differences within the 3 good-clone group ($G - 2$ d.f.)
(iii) Differences within the 3 poor-clone group ($P - 2$ d.f.)

I have adopted a similar partition in clone $\times$ block interactions (Error 1), in $C \times T$ and in $C \times T \times$ Blocks (Error 2), with the idea of testing C-components for significance against their corresponding C components $\times$ block interactions.

On carrying out the tests of significance in the covariance analysis, Table 2, I find C, when compared with BC, to be significant at the 0.1% point but when comparing C' with $C'B$, or G with GB, or P with PB, no significant effects emerge, although I should expect C' to be even more highly significant than C! When C' is compared with BC, however, it is found to be highly significant, yet BC and BC' do not differ significantly.

TABLE 2
ANALYSIS OF COVARIANCE IN SPLIT PLOT EXPERIMENT

Source of Variation	D.F.	Sx^2	Sxy	Sy^2
Blocks (B)	5	18.79	13.63	14.19
Clones (C)	5	186.14	221.12	340.96
Good v Poor (C')	1	124.79	185.47	275.68
Within Good (G)	2	27.34	34.70	64.16
Within Poor (P)	2	34.01	0.95	1.12
Error 1 (BC)	25	47.31	12.24	74.39
$C'B$	5	2.09	2.06	7.74
GB	10	9.44	4.07	54.69
PB	10	35.78	6.11	11.96
Buds v Cuttings (T)	1	141.91	6.91	0.34
CT	5	72.43	12.48	22.47
$C'T$	1	18.03	14.56	11.75
GT	2	24.71	5.09	2.28
PT	2	29.69	−7.17	8.44
Error 2	30	109.35	34.09	54.07

Am I wrong in attempting to compare treatment components such as C', G and P with their interactions with blocks, or should BC be used in tests of significance? If the former procedure is correct, how does it come about that none of the C-components attains significance, especially as no significant difference exists between BC and BC'.

Is it possible that mere linear correction is inadequate in this particular case and that a more valid analysis of covariance could involve the squares of ages?

ANSWER: First, a preliminary point not inquired about: if the clones were selected for trial because half were good and half poor, then the testing of C' is justified by the design of the experiment; but if you designated the clones as good and poor because of their performance during the experiment, then I am not happy about the test. See this Journal, Vol. 1: page 26; Vol. 2: page 16; and Vol. 5: page 99. Since the trees (or the experiment) are 8 years old, and since I have observed no mention of good and poor clones before 1945, I wonder if you are justified in the testing of C'. Actually, the yield of Clone 44 differs more from that of Clone 1 than it does from the yields of the "poor" clones.

Granting that the design of the experiment included the comparison of the three good clones with the three poor, I think you are correct in testing C', G and P against their discrepancies with blocks. Each discrepance is expected to estimate the real experimental error to which the corresponding effect is liable. BC and BC' are not orthogonal, so that test gives little information.

The arithmetical peculiarities which puzzle you are easily explained by plotting the mean yields against the ages, then drawing the error regression line. As an example, the line connecting the points representing the means for the good and poor clones will be found to be sensibly parallel to the error regression $C'B$, the slopes being $185/125 = 1.5$ and $2.06/2.09 = 1.0$. So a large fraction of the difference in yield, C', is attributable to the age difference between the plantings. This explains the non-significance of C'.

I doubt if curved regression is responsible for the irregularities in your data. Rather, I suspect that you have gained little advantage from the use of age as a covariate. There are two reasons. The first lies in the possible confusion of ages with yield among the 8 trees of each plot. In some plots, the average age is little more than half that in others. Either all the trees in these plots have been replaced by younger, or else some of the replacements have scarcely come into bearing. Unless the yield-growth curves of all these ages are straight and identical, the averages do not give an accurate measure of the yield which is associated with any specific age. As a preliminary, I should like to have a look at the yields of all the trees that have a common age. This would give me clearer information as to whether differences among the clones persist among trees of given age.

This brings me to the second reason for questioning the use of age as a covariate. Perhaps age is itself an informative measure of the value of the clones. The data seem to indicate that some clones are adapted to the environment and others not. In both breeding and selection, livability as well as yield would seem to demand consideration.

78 QUERY: I have seen the statement offered without proof that the most economical test of the difference between two unmatched groups is achieved when the numbers of observations in the two groups are equal. Thus, let us suppose that an experimenter wants to test whether two different methods of learning are equally good. He has two groups of subjects, A and B, unmatched in any way, and has both groups learn the same material. Group A learns by one method; B by the other. Assume that the two methods really do give different results. In general, will this experimenter be able to discover the difference with the smallest total number of tests if the number of subjects in group A equals that in group B? If this is true, is there some source which gives the proof for this statement?

ANSWER: Suppose we are trying to measure the difference between two quantities as precisely as possible with a fixed amount of effort. By "precisely" we mean that we should like the variance of our estimate of the difference to be as small as possible. But, if the measurements are independent, the variance of the difference of the means is the sum of the variances of the two means:

$$V(\bar{x} - \bar{y}) = V(\bar{x}) + V(\bar{y}).$$

Further, we know that

$$V(\bar{x}) = V(x)/n_x ,$$

$$V(\bar{y}) = V(y)/n_y ,$$

where $V(x)$, $V(y)$ are the variances of the x and y distributions. We therefore want to minimize

$$\frac{V(x)}{n_x} + \frac{V(y)}{n_y}$$

subject to the condition that

$$n_x + n_y = n$$

where n is fixed. It is not difficult to show that the minimization occurs for

$$n_x/n_y = \sigma_x/\sigma_y ,$$

that is, that the numbers of observations should be in the ratio of the standard deviations. In the special case where the standard deviations are equal, the observations should be equally divided.

That the distribution of effort should depend on the variability of each group is clear if we consider the case where one variance is zero, for plainly a single observation will suffice for this group.

In practice, of course, we usually do not know the variances beforehand, and so cannot use the result directly. But also, if we do not know anything about the variances, it seems sensible to make the numbers equal.

I am afraid I cannot give you a reference to a proof, though I have no doubt that the result is old. Somewhat similar problems have been considered in connection with sample surveys.

C. P. WINSOR

79 QUERY: The problem upon which I am engaged is the determination of the relative effects of various factors upon yield and size of orange fruits. Of course, yield and size are inversely correlated. Our main endeavor is to determine whether or not yield or size is correlated with the quantity of nutrients absorbed. For this purpose it has seemed probable that the use of standard partial regressions would be profitable and in fact this is turning out to be the case.

However, a question has arisen which I shall appreciate having answered. I have assumed that a standard partial regression coefficient would have a value no more than 1. However, in one of our problems in multiple regression we obtain a standard partial regression coefficient of 1.3. I have looked through a number of texts and through the data and computations for a considerable number of partial regression problems which I ran some years ago, and in none of these instances have I happened to find a value greater than 1. I will be grateful for a statement as to the theoretical possibility of obtaining a value greater than 1.

ANSWER: There is no arithmetical limitation on the size of a standard partial regression coefficient. My experience has coincided with yours in that, for the usual type of linear regression, these coefficients tend to be small. If I get one larger than 2, I begin to look for trouble; but larger values can occur.

CORRECTION IN QUERY 74, DECEMBER, 1949. Dr. Tukey writes that Professor Royal F. Bloom has pointed out an error in calculating chi-square for the difference,

$$|\text{observed} - \text{expected}| - 1/2.$$

The difference should be 48.8, with the correct value of chi-square, 41.2. Since this is for 1 degree of freedom, its square root, 6.41, is roughly a normal deviate and compares adequately with the value, 6.13, obtained by the more direct method.

ABSTRACTS

EASTERN NORTH AMERICAN REGION ANNUAL MEETING OF THE BIOMETRIC SOCIETY

New York, December 28–30, 1949

LURIA, S. E. and R. DULBECCO (University of Indiana).

87 **Interpretation of the Formation of Active Bacterial Virus From Ultraviolet Inactivated Virus** (*Genetics* 34: 93–125).

Bacteriophage particles are inactivated at a logarithmic rate by ultraviolet light. The ln of the survival ratio gives the average number of inactivating hits (r) per particle. When two or more inactive particles of the same bacteriophage are adsorbed by a bacterial cell, there is a production of active bacteriophage in a fraction of these cells. This fraction increases with the average number x of particles adsorbed per bacterium and decreases with increasing r. These results are interpreted as due to the production of lethal mutations by ultraviolet in discrete genetic units, of which each particle contains a constant number n. From the experimental values of r, x, and of the probability that a bacterium liberates active phage, a relation is derived between these parameters, on the assumption that active virus is produced in a bacterium whenever the infecting particles as a group contain at least one copy of each of the n units in nonlethal form. The theoretical relation fits the experimental results rather closely and leads to a calculation of a minimum value for n for each bacteriophage. The limitations of this analysis and some systematic deviations from experiment were pointed out and discussed.

NEWMAN, E. V.; M. MERRILL. (Johns Hopkins University).

88 **The Application of Equations Derived from Models, to 'Central' Circulatory Volume.**

The purposes of our studies are to (1) derive a theory which expresses the concentration change in the outflow fluid of a flow system such as the human heart and lungs after a single instantaneous injection of a known amount of an indicator substance such as T-1824, (2) test the theory by comparison to dilution curves obtained from mechanical models in which the flow and the volumes of the compartments are known, and (3) apply the theory to the analysis of dilution curves obtained with human subjects.

An equation was derived which expresses the variation of concentration with time as a function of the amount of dye injected, the rate of flow through the system, and the volumes of three chambers in series. The right heart, the lungs and the left heart are theoretically considered as the three separate successive volumes in which the dye is mixed and diluted.

The dilution curves obtained from a mechanical model are nearly identical with the theoretically derived curves. The equation gives the outflow-fluid concentration as the algebraic sum of three exponentials whose rate constants are made up of known constants consisting of the volumes in the system, the amount of dye injected, and the flow through the system.

Comparison of the theoretically derived and mechanically produced dilution curves with human curves shows close similarity. The relationship of the constants derived from human curves to the volume of blood in the heart and lungs is discussed. These mathematical and mechanical models provide a basis for the rational interpretation of human dye-dilution curves.

A device for rapid accurate collection of serial samples from flow systems has been constructed.

89 DENSEN, P. M. (University of Pittsburgh). **A Definition of the Group to be Followed** (To appear in *Human Biology*).

Follow-up studies in relation to morbidity of one kind or another have been appearing in the literature with increasing frequency as a direct consequence of the growing importance of the chronic diseases as causes of morbidity and mortality. A clear definition of the group to be followed is a necessary prerequisite to such studies because the objective is to obtain a set of facts which may serve as the basis for predictions about other similar cases. It is only to the degree that the study cases have been precisely defined that this similar group can be adequately described and the extent indicated to which generalization may be permitted. Several examples are given to illustrate this point.

Definition from the standpoint of diagnosis is only a part of the total definition. Other factors besides diagnosis, such as age, sex, duration of infection prior to diagnosis, etc., may play a very important part in the determination of the universe to which the findings may be generalized.

Often an investigator may have difficulty in finding material to work with of precisely the kind he needs and he may decide to take the best available. This is fair enough as long as he recognizes that the

universe to which generalization is made may not be the universe he originally had in mind. It is essential that the investigator specify the nature of the selection in order that others may make proper use of the findings. In this connection it must be recognized that a definition can be adequate only in relation to some specific objective. It is part of the job of the statistician to insist that a definition be made and that its implications be understood. It is not his job, however, to make the definition.

90 HARRIS, T. E. (Rand Corporation), PAUL MEIER (Philadelphia Tuberculosis and Health Association), and JOHN W. TUKEY (Princeton University). **Timing of The Distribution of Events Between Observations.***

When all that is known as to the time of occurrence of an event is that it had not occurred at one known time ("last previous") and that it had occurred at another known time ("first after"), it is ordinarily useless to try to date the event more closely. But when we have information of this nature about many events, and when it is reasonable to think of the events as a sample from a distribution, it is possible to use the information to estimate the distribution. It is natural to express this distribution in terms of event-rates. The problem discussed here is how these event-rates may reasonably be estimated from such observations.

Any attempt to infer the timing of events from sparse observations is subject to many pitfalls, some of which are discussed briefly.

If a given body of data has avoided these pitfalls, a simple statistical model may apply, for which the best standard estimation procedures are available. The resulting maximum likelihood estimates are shown to correspond quite closely to those found by a very simple routine, namely:

(1) Each case between "last previous" and "first after" contribute one half to the number of cases exposed to risk,

(2) Each case's event is distributed over the interval between "last previous" and "first after" according to the finally estimated chance of occurrence.

(The appearance of the one-half and of these rules is a matter of mathematics, rather than the result of simplifying assumptions.)

An iterative procedure of obtaining exact maximum likelihood estimates, and making rough (but usually adequate) significance tests is given, and the danger that maximum-likelihood-like estimates may follow the irregularities of the data too faithfully is discussed.

---

*Prepared in connection with research sponsored by the Office of Naval Research.

These procedures, and some modifications, are illustrated on a simple numerical example and on the "progression" of the cases of minimal tuberculosis. The latter example is based on the records of the Henry Phipps Institute as studied by Drs. A. L. Cochrane, H. W. Campbell and S. C. Stein. The present analysis confirms their conclusions, as obtained by simpler methods, and exhibits a marked tendency for the progression dates to avoid a period of 9 to 12 months after detection. The possible significance of this result is discussed.

91 DORN, HAROLD F. (National Institute of Health). **Methods of Analysis for Follow-Up Studies.**

Two practical difficulties arise in long time follow-up morbidity studies. (a) The impossibility, without excessive cost, of keeping each member of a group of cases under observation, resulting in only partially complete information, and (b) the necessity of combining, in a meaningful way, the experience of cases of varying durations. Three concepts essential in the analysis of data from such studies are defined: (a) cohort of cases, (b) person-time units of exposure and (c) exposure to risk. Principles and procedures pertaining to the following problems are discussed: (a) cases to be included, (b) starting date, (c) classification of status, (d) the length of the interval of tabulation, (e) time at which the status of an individual is to be determined, (f) handling of cases observed for part of an interval, and (g) handling of cases lost to follow up. Formulas for different types of rates are given and the arrangement and construction of a morbidity table is illustrated.

92 WORCESTER, JANE (Harvard University) and STUART S. STEVENSON (University of Pittsburgh). **Malformations in the Boston Lying-In Hospital, 1930-1941.** (To be published in *Pediatrics*).

The gestational characteristics associated with the birth of 677 congenitally malformed infants have been studied. The difficulties involved in the selection of control groups are discussed.

93 YOUDEN, W. J. (National Bureau of Standards). **Index for Rating Diagnostic Tests.**

Diagnostic tests have the task of correctly designating which are the diseased individuals in a population. The test may err in giving negative tests for diseased individuals (false negatives) and in giving positive tests for healthy individuals (false positives). An index,

$$J = \frac{ad - bc}{(a + b)(c + d)},$$

is proposed as a measure of performance of a diagnostic test, where a is the number correctly diagnosed out of $a + b$ diseased individuals, and d is the number correctly reported negative out of $c + d$ controls. The index provides a ready means of comparing the merits of alternative diagnostic tests.

94 GREENHOUSE, SAMUEL W. and NATHAN MANTEL (National Cancer Institute). **The Evaluation of Diagnostic Tests.**

This paper presents statistical procedures for evaluating and comparing diagnostic tests which yield a continuous range of scores for the known positive and negative individuals tested. These procedures are developed both assuming normality of the distribution of diagnostic test scores, and making no assumption about distribution forms. The evaluation procedures shown do not make use of any critical point which distinguishes between positives and negatives. This makes it possible for the statistician to make an evaluation unhampered by any prior judgements of the value of the critical point.

For a single diagnostic test, the procedure for determining sample size necessary for evaluation is given. Criteria for selection of a critical point for differentiating between positives and negatives for a single test are also considered.

In the comparison of two diagnostic tests, procedures are developed for the case where the same individuals are used for both tests, and also where different individuals are used for each of the tests.

FIRST INDIAN REGION BIOMETRIC CONFERENCE

Poona, January 1950

95 DANDEKAR, V. M. **Certain Modified Forms of Binomial and Poisson Distributions.**

The Binomial and Poisson distributions in which the probabilities at successive trials are independent have been slightly modified by introducing the condition that the probability of occurrence is zero for a number of trials immediately following an occurrence. Appropriate distributions have been obtained and the occasions suitable for their application indicated.

96 BANERJEE, BASUDEB and ANUKUL CHANDRA DAS. **On the Response of Mimose-Pudica Leaflets to its Organ Extract.**

Ricca's observation, that leaflets of *Mimose-Pudica* respond to a Chemical substance and it is present in its organ extract, led the botanists to isolate that Chemical substance. A purer sample of the extract will give response to a higher dilution.

In this paper it has been established that for determining response, time at the beginning of reaction should be noted and time for its completion is not needed. The latter is constant for all concentrations and under all atmospheric conditions but the former varies. But whether the latter is constant for all degrees of purity is to be investigated further.

97 RAO, S. RAJA. **Normal Curve as an Approximation to Statistical Distributions.**

The probability integral tables of Statistical Distributions are used by Statisticians mostly in tests of significance. For this purpose all that we need to know is the value of the "Statistic" at specified percentage points like .5 per cent, 1 per cent, 2.5 per cent, 5 per cent, etc.

In this note, combinations of β_1 and β_2 have been obtained such that

$$\left| \int_{x_1}^{x_2} F(x)\, dx \right| < 0.0025 \qquad \text{where } x_1 \text{ and } x_2 \text{ are such that}$$

$$\int_{\infty}^{x_1} F(x)\, dx = \int_{-\infty}^{x_2} \phi(x)\, dx = \alpha \text{ (a given probability) and}$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

and $F(x)$ is the Gram-charlier series of type A with the first five terms which can be taken as a good representation of moderately skew Distributions. Under these conditions x_2 is a good approximation to x_1 .

Suitable graphs have been drawn to facilitate the use of this method in such approximations. Making use of this criterion, the size of the sample beyond which the sampling distributions can be considered as normal for purposes of tests of significance has been worked out for some important Statistical distributions like Student's t, x^2, etc.

98 POTI, S. JANARDAN, **Design and Analysis of Blood Group Sample Surveys.**

An attempt has been made to fix the sample size for detecting with

confidence any given differences in the gene frequencies of two groups of individuals. Tables for some typical values of the gene frequencies have been calculated.

99 RAO, C. RADHAKRISHNA, **The Distribution of the Difference Between the D^2 Statistics Based on p and p + q Characters.**

In a paper published in Sankhya, Vol. 9, Part 4, the author derived the distribution of $D^2_{p+q} - D^2_p$ when the population value Δ^2_p is zero. In this paper the distribution is derived for nonnull Δ^2_p and some illustrations have been given when the statistic based on the difference $D^2_{p+q} - D^2_p$ is useful in tests of significance.

ANNUAL MEETING OF THE BRITISH REGION OF THE BIOMETRIC SOCIETY

THE MEDICAL SOCIETY OF LONDON, MARCH 14, 1950

The Medical School of London

March 14, 1950

100 GRIDGEMAN, N. T. **The Graphical Calculation of the Results of Biological Assays with Graded Responses.**

The fiducial limits of error, at a given probability level, of a balanced assay (response linearly related to log dose) can be expressed as

$$I[R(C-1) \pm \sqrt{(C-1)(R^2C+1)}]$$

where I = log dose-interval, R = standard-test response difference as a fraction of dose-interval response difference, and C = Fieller's correction factor. A chart has been constructed enabling percentage limits of error (100 × antilogs of the above expression) to be read off.

101 WOOD, E. C. **The Estimation of Error in Certain Types of Biological Assays.**

In those biological assays in which one animal from each litter is assigned to each treatment-group, the 'residual' component of variance after removing the Treatment and Litter components is used for estimating the error mean square. If, however, two or more litter-mates can be assigned to each treatment-group, as occasionally becomes possible, an estimate of the true error mean square is available independently of the Treatment × Litter interaction. It is then sometimes found that the

latter, or some of the orthogonal components into which it can be partitioned, is significantly greater than the error. The meaning of this finding, and its effect on the calculation of the fiducial limits of the result, are discussed.

102 FIELLER, E. C. **The Problem of Combining the Results of Independent Assay.**

The paper proposes methods for combining the data of independent assays with a continuous response:

(i) When the slope and residual variance are stable from one assay to another,

and

(ii) When the slope varies while the residual variance remains constant,

and discusses the application of these methods to litter-mate assays.

JOINT MEETING OF THE INSTITUTE OF MATHEMATICAL STATISTICS AND THE EASTERN NORTH AMERICAN REGION OF THE BIOMETRIC SOCIETY

Chapel Hill, North Carolina

March 17–18, 1950

103 GHURYE, S. G. (University of North Carolina). **A Method of Estimating the Parameters of an Autoregressive Time Series.**

The general autoregressive process of the second order is defined by the equations

$$x_t = X_t + \eta_t$$

and

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} = \epsilon_t \ ,$$

where x_t is the value actually observed at time t, X_t the corresponding theoretical value, ϵ_t the disturbance and η_t the superposed variation. The estimates of α_1 , α_2 given by Yule's method are biased and inconsistent if η_t is not identically zero, the permanent bias being a function of the unknown variance of η_t . The present paper proposes a method of estimation which is unaffected by the presence of η_t , and seems to be better than any other known method; and this conjecture is supported

by the results of application to observational and artificial series. In this method, the estimates a_1 , a_2 are obtained by minimizing

$$\sum_{k=3}^{n} \frac{1}{(N-k-2)} \cdot \left\{ \sum_{t=3}^{N-k} (x_t + a_1 x_{t-1} + a_2 x_{t-2})(x_{t+k} + a_1 x_{t+k-1} + a_2 x_{t+k-2}) \right\}^2$$

where n is some number small in comparison with N (which is the number of observations). In the above expression, the usual approximation of substituting

$$(N-k-2)r_k \qquad \text{for} \qquad \sum_{t=3}^{N-k} x_t x_{t+k}$$

may be made for computational convenience. The method has been used for fitting autoregressive processes to the series of annual averages of Wolfer's sunspot numbers and that of Myrdal's Swedish cost of living index numbers. The method is applicable to higher order processes.

104 HOEFFDING, WASSILY. (Institute of Statistics, University of North Carolina). **Most Powerful Rank Order Tests.**

Let X_{11} , $\cdots$, X_{1n_1} , $\cdots$, X_{k1} , $\cdots$, X_{kn_k} be random variables with a joint probability function $P(S)$, and let $P\{X_{iu} = X_{jh}\} = 0$ if $g \neq h (i = 1, \cdots, k)$. Let H_0 be a hypothesis which implies that $P(S)$ is invariant under all permutations of X_{i1} , $\cdots$, $X_{in_i} (i = 1, \cdots, k)$. Let $r_{ij} (j = 1, \cdots, n_i)$ be the ranks of X_{i1} , $\cdots$, X_{in_i} . Under H_0 the $M = \Pi n_i$: rank permutations $R = (r_{11}, \cdots, r_{1n_1}, \cdots, r_{k1}, \cdots, r_{kn_k})$ have the same probability $P(R) = M^{-1}$. A test which depends only on the permutations R is called a rank order test (R.O.T.). A R.O.T. of size m/M which is most powerful (M.P.) against a simple alternative, $P_1(S)$, is determined by m permutations R for which $P_1(R)$ takes on its m largest values.

For example, let the pairs (X_1, Y_1), $\cdots$, (X_n, Y_n) be independent and identically distributed. Let H_0 state that X_i , Y_i are independent, and let $H_1(\rho)$ be the hypothesis that X_i , Y_i have a bivariate normal distribution with correlation ρ. We may assume that $X_1 < \cdots < X_n$ and consider the ranks r_i of the Y's only. A R.O.T. which is uniformly M.P. against all $H_1(\rho)$ with $\rho > 0$ does not exist except for small n. The M.P.R.O.T. against small $\rho > 0$ is determined by the largest values of $\sum_{i=1}^{n} (EZ_i)(EZ_{r_i})$, where EZ_i is the expectation of the i-th order statistic in a sample of n from a standard normal distribution. The M.P. unbiased R.O.T. against small values of $|\rho|$ is based on the statistic $\sum_i \sum_j (EZ_iZ_j)(EZ_{r_i}Z_{r_j})$. The M.P.R.O.T.

against ρ close to 1 is obtained by expanding the probability of R in powers of $\{(1 - \rho)/(1 + \rho)\}^{1/2}$.

COCHRAN, WILLIAM G. (Department of Biostatistics, Johns Hopkins University). **105 The Comparison of Percentages in Matched Samples.**

In this paper the familiar χ^2 test for comparing the percentages of successes in a number of independent samples is extended to the situation in which each member of any sample is matched in some way with a member of every other sample. This problem has been encountered in the fields of psychology, pharmacology, bacteriology, and sample survey design. A solution has been given by McNemar (1949) when there are only two samples.

In the more general case, the data are arranged in a two-way table with r rows and c columns, in which each column represents a sample and each row a matched group. The test criterion proposed is

$$Q = \frac{c(c - 1) \sum (T_j - \overline{T})^2}{c(\sum u_i) - (\sum u_i^2)}$$

where T_j is the total number of successes in the j-th sample and u_i the total number of successes in the i-th row. If the true probability of success is the same in all samples, the limiting distribution of Q, when the number of rows is large, is the χ^2 distribution with $(c - 1)$ degrees of freedom. The relation between this test and the ordinary χ^2 test, valid when samples are independent, is discussed.

In small samples the exact distribution of Q can be constructed by regarding the row totals as fixed, and by assuming that on the null hypothesis every column is equally likely to obtain one of the successes in a row. This exact distribution is worked out for eight examples in order to test the accuracy of the χ^2 approximation to the distribution of Q in small samples. The number of samples ranged from $c = 3$ to $c = 5$. The average error in the estimation of a significance probability was about 14 per cent in the neighborhood of the 5 per cent level and about 21 per cent in the neighborhood of the 1 per cent level. Correction for continuity did not improve the accuracy of the approximation, although it is recommended when there are only two samples. Another approximation, obtained by scoring each success as "1" and each failure as "0" and performing an analysis of variance on the data, was also investigated. The F-test, corrected for continuity, performed about as well as the χ^2 approximation (uncorrected), but is slightly more laborious.

The problem of subdividing χ^2 into components for more detailed tests is briefly discussed.

106 LUCAS, H. L. (North Carolina State College). **A Method of Estimating Components of Variance in Disproportionate Sub-Class Data.**

By including sufficient effects in the forward solution of the Abbreviated Doolittle method, components of variance may be estimated from disproportionate data. The procedure is very systematic, and thus, is adaptable to routine computational work. The computations will be described, and the utility of the method briefly discussed.

107 ISAACSON, STANLEY L. (Columbia University). **On the Theory of Unbiassed Tests of Simple Statistical Hypotheses Specifying the Values of Two Parameters (Preliminary Report).**

In the Neyman-Pearson theory of testing simple hypotheses, in the one-parameter case, a locally best unbiassed region is called "type A." It is obtained by maximizing the curvature of the power curve at the point $\theta = \theta_0$ specified by the hypothesis, subject to the conditions of size and unbiassedness. For the two-parameter case, Neyman and Pearson considered "type C'" regions (*Stat. Res. Mem.*, vol. 2 (1938), pp. 36, ff.). The definition of these regions requires one to choose in advance a family of ellipses of constant power in an infinitesimal neighborhood of the point $(\theta_1 , \theta_2) = (\theta_1^0 , \theta_2^0)$ specified by the hypothesis. The natural generalization of a "type A" region is a "type D" region, which maximizes the Gaussian curvature of the power surface at $(\theta_1^0 , \theta_2^0)$, subject to the conditions of size and unbiassedness. This definition does not require one to choose a family of ellipses in advance. This approach leads to a new problem in the calculus of variations. A sufficient condition is obtained which plays the role of the Neyman-Pearson fundamental lemma in the "type A" case. An illustrative example is given. (Prepared under sponsorship of the Office of Naval Research.)

108 BOSE, RAJ CHANDRA. (Institute of Statistics, University of North Carolina). **A Note on Orthogonal Arrays.**

Consider a matrix $A = (a_{ij})$ with N rows and m columns, each element a_{ij} standing for one of the $s - 1$ integers $0, 1, 2, \cdots , s - 1$. Let us take the partial matrix obtained by choosing any $t < m$ columns of A. Each row now consists of an ordered t-plet of numbers, and each element has one of s possible values, there are s^t possible t-plets. The matrix A may be called an orthogonal array (N, m, s, t) of size N, m constraints, s levels and strength t, if by choosing any t columns whatsoever every possible t-plet occurs the same number of times. Clearly $N = \lambda s^t$ where λ is an integer. Such arrays have been considered by

Rao and are useful for various experimental designs. The existence of an orthogonal array $(s^2, m, s, 2)$ is equivalent to the existence of a set of orthogonal Latin squares of side s and m constraints (i.e. the number of Latin squares in the set is $m - 2$). The fundamental question that can be asked regarding orthogonal arrays is the following: What is the maximum numbers of constraints for an orthogonal array, given N, s and t? Denote this number by $f(N, s, t)$, then from known properties of Latin squares $f(s^2, s, 2) = s + 1$, if s is a prime or a prime power, and a theorem by Mann states that $f(s^2, s, 2) \geq r + 1$, if $s = p_1^{n_1} \cdots p_k^{n_k}$ where $p_1, \cdots, p_k$ are different primes, and r is the minimum of $p_0^{n_0}, p_1^{n_1}, \cdots p._k^{n_k}$ The following generalisation of Mann's theorem is proved in this note.

$$f(N_1N_2 \cdots N_k, s_1s_2 \cdots s_k, t) \geq \text{Min } f(N_1, s_1, t), f(N_2, s_2, t),$$

$$\cdots, f(N_k, s_k, t)$$

109 FREEMAN, MURRAY F. and JOHN W. TUKEY (Princeton University). **Transformations Related to the Angular and the Square Root.***

The use of transformations to stabilize the variance of binomial or Poisson data is familiar (Anscombe, Bartlett, Curtiss, Eisenhart). The comparison of transformed binomial or Poisson data with percentage points of the normal distribution to make approximate significance tests or to set approximate confidence intervals is less familiar. Mosteller and Tukey have recently made a graphical application of a transformation related to the square-root transformation for such purposes, where the use of "binomial probability paper" avoids all computation. We report here on an empirical study of a number of approximations, some intended for significance and confidence work, and others for variance stabilization.

110 VERLINDEN, F. J. (North Carolina State College). **Standard Inverse Matrices for Fitting Polynomials.**

For fitting polynomials of the type, $y = b_0x^0 + b_1x + b_2x^2 + \cdots + b_mx^m$, with the x's equally spaced, published tables of orthogonal polynomials may be used. This procedure does not yield the b's directly, nor their variances or covariances, although such may be obtained by proper computations which are moderately tedius. In some types of statistical work, the b's and their variances and covariances may be desired. These may of course be obtained directly by the method of least squares but the computational work is prodigious relative to that

*Prepared in connection with research sponsored by the Office of Naval Research.

for the orthogonal polynomial approach. When the x's are equally spaced the elements of the variance-covariance matrix may be put in the simple form of sums of powers (including the zero power) of successive integers from zero to n (n equals one less than the number of observations). The elements of the inverses of matrices of this type have been worked out algebraically in terms of n for polynomials up to and including the quintic ($m = 5$). With these standard inverse matrices, the b's and their variances and covariances may quickly be obtained once the elements are evaluated numerically. These elements have been evaluated numerically up to $n = 20$.

111 RAFFERTY, J. A., M.D. (Dept. of Biometrics, School of Aviation Medicine, Randolph Field, Texas). **Mathematical Models in Biology.**

From the point of view of a bio-medical research administrator, mathematical models will assume a greater role in biological research than heretofore. In anticipation of this trend, certain philosophical implications of models in biological theory and scientific theory in history are examined. A hierarchy of abstraction-levels in biology is delineated, and the role of mathematical models at these levels is illustrated by examples from the literature. Proposals are made for a concentration of mathematical effort on certain important biological problems. Remarks are made on the capabilities and limitations of models in biology.

112 BROSS, IRWIN. (Johns Hopkins University). **Small Sample Performance of Biological Statistics.**

In this paper the dilution method for estimating bacterial density is investigated by an exact small sample method and also by an approximate one. Methodologies and design of experiments are compared for various small sample cases.

113 GREENBERG, B. G. and A. HUGHES BRYAN. (University of North Carolina). **Methodology in the Study of Physical Measurements of School Children.**

In a series of investigations to determine by small-sampling technique what physical differences, if any, occur between children of differing socio-economic backgrounds, several problems of methodology arose. A pilot study was undertaken to assure maximum efficiency at each step. It was found that the children could remain dressed (with the exception of boys' bi-iliac measurement) without changing the magnitude of the

differences. The pilot study enabled us to decide how many observers to use, and how much duplication of measurements by them was necessary. Minimum sample sizes were estimated to indicate physical differences of predetermined magnitudes. It was found that the age grouping 96–143 months was optimal from the standpoint of indicating physical differences between children of differing socio-economic levels. Boys and girls in the upper socio-economic levels were both taller and heavier for their age in this age group. There were no weight differences, however, when weight was adjusted for age and height. Measurement of the bi-iliac and transverse chest diameter provided little additional information on physical differences. The calf circumference, an indicator of muscle mass and subcutaneous fat, is suggested as being a sensitive supplementary index to indicate physical differences when age and height are adjusted.

114 HOUSEHOLDER, A. S. (Oak Ridge National Laboratory). **Tetrad Analysis in Yeast.**

In neurospora all four products of meiosis are recovered in the four spores of an ascus. In crosses $AB \times ab$ the asci are of three types, designated I, II or III according as all four, none, or two spores resemble parents. Frequencies of these types, P, P' and P'' are the observables. If there were no exchange P'' would be zero; and one should have $P' = 0$ or $1/2$ according to whether the loci were on the same or different chromosomes.

Assuming only that no exchange occurs between sister chromatids and neglecting chromatid interference, one can calculate without further assumptions a frequency P'' of exchanges between a single locus and its centromere from data on three or more genes taken in pairs by equations

$$s_{ij} = s_{0i}s_{0j}\ , \qquad P'' = 2(1 - s)/3,$$

where the subscript 0 refers to a centromere. Lindegren makes such calculations from his own data, by taking groups of three, but makes no effort to reconcile discrepancies. Neyman's modified chi-square, however, permits combining all observations in a set of equations that yields easily to a rapidly converging iterative solution. The equations are

$$2s_i \sum_{j \neq i} s_j^2(n_{ij} + n'_{ij})^2(n_{ij}^{-1} + n_{ij}'^{-1}) = \sum_{j \neq i} s_j(n_{ij} + n'_{ij})^2(2n_{ij}^{-1} - n_{ij}'^{-1}),$$

where n_{ij} is the number in class I and II combined for the loci i and j, n'_{ij} the number in class III, and only those pairs (i, j) are included which are found to be independent.

The argument of A. R. G. Owen (Pr. R. S. B. 136, '49), p. 67 can be

paraphrased for the present case and a suitable generating function $P(\lambda, u)$ is being sought providing a metric. The specific one proposed by Owen is ruled out since

$$s = P(-\tfrac{1}{2}, u)$$

takes on a negative value for one locus, which is not possible with Owen's function.

115 RAPOPORT, ANATOL. (University of Chicago). **Contribution to the Probabilistic Theory of Neural Nets. I. Randomization of Refractory Periods and of Stimulus Intervals.**

Aggregates of neurons are considered in which the frequency of occurrence of neurons with a specified value of the refractory period follows certain probability distributions. Input-output functions are derived from such aggregates. In particular, if input and output intensities are defined in terms of stimulus frequencies and firing frequencies per neuron respectively, it is shown that a rectangular distribution of refractory periods leads to a logarithmic input-output curve. If input and output are defined in terms of the total number of stimuli and firings in the aggregate, it is shown how the "mobilization" picture leads to the logarithmic input-output curve.

By randomizing the intervals between stimuli received by a single neuron and by introducing an inhibitory neuron a very simple "filter net" can be constructed whose output will be sensitive to a particular range of the input, and this range can be made arbitrarily small.

116 LANDAHL, H. D. (University of Chicago). **Theoretical and Experimental Aspects in the Removal of Air-Borne Matter by the Human Respiratory Tract.**

The principal factors governing the fate of a particle in the respitatory tract are impaction due to inertia, settling due to gravity and Brownian movements. For a given respiratory pattern, it is possible to calculate the probable fate of a particle from a knowledge of the geometry of the passages. These calculations have been carried out in such a manner as to obtain the theoretical amounts of material deposited in various regions of the lungs as well as the relative amounts in various fractions of the expired air. Similarly, it is possible to estimate the probable fate of a particle which passes through the nasal passages. Experiments have been carried out to verify a number of these predictions. On the whole, the agreement, as illustrated in the slides, is fairly satisfactory when one considers the complexity of the calculations.

117 WADLEY, F. M. (Navy Department). **An Application of Biometrics to Zoological Classification.**

Statistical problems in taxonomy are discussed; attention must be paid to variation of individuals as well as of group means. Covariance analysis and the discriminant function technique are applied to multiple measurements in groups of molluscan fossils.

118 MOSHMAN, JACK. (United States Atomic Energy Commission). **The Analysis of Hemotological Effects of Chronic Low-Level Irradiation.**

Several methods are investigated for analyzing the possible effects of chronic low-level irradiation upon the employees of the operating contractors of the US AEC. The effects investigated are those on the red blood count, hemoglobin, white blood count, lymphocytes and neutrophils. The analysis includes measurements of significant differences among individuals, geographic sites and the exploration of various indices of exposure to radiation. A non-parametric determination of trend values for individuals which may be applied to mass data is considered.

119 CURETON, EDWARD E. (University of Tennessee). **Statistical Problems in Psychological Testing.**

Though great progress has been made in mathematical statistics in recent years, a number of the major statistical problems encountered in the development and use of psychological tests remain unsolved. Some of these problems are outlined, with particular reference to the mathematical models and assumptions implied by psychological theory, by the nature of the experimental data, and by the conditions under which the results and findings are to be applied.

120 NICHOLSON, GEORGE E., JR. (Institute of Statistics, University of North Carolina). **Accuracy of a Linear Prediction Equation in a New Sample.**

The problem considered is as follows. Given two samples S_1 and S_2 of N_1 and N_2 observations on a $p + 1$ character random variable $(y, x_1, \cdots, x_p)$. Let Y_1 and Y_2 be the linear regression equation computed by the method of least squares from each sample. The effect of using Y_1 to predict the y's in S_2 is considered. The ratio

$$k \cdot \frac{S(y_2 - Y_1)^2}{S(y_2 - Y_2)}$$

is used as a measure of the predicting efficiency of Y_1 in S_2 relative to Y_2 when the X_i are fixed for the usual regression model. The general multivariate case is also considered.

121 BAHADUR, RAGHU RAJ. (Institute of Statistics, University of North Carolina). **Smallest Average Confidence Sets for the Simultaneous Estimation of K Normal Means.**

Let $v = (x_{11}, \cdots, x_{1n1}; \cdots; x_{k1}, \cdots, x_{knk})$ denote the combined sample point in samples of sizes $n_1, n_2, \cdots, n_k$ from normal populations $\pi_1, \pi_2, \cdots, \pi_k$ respectively, π_i having mean μ_i and variance σ_i^2. Writing $\mu = (\mu_1, \mu_2, \cdots, \mu_k)$, denote the k dimensional Euclidean space of all points μ by R. Given any parameter point (μ, σ), where $\sigma = (\sigma_1, \sigma_2, \cdots, \sigma_k)$, and any set valued function $f(v)$ defined for all sample points v and having subsets of R as its values (which satisfies certain measurability hypotheses), let $\alpha(f \mid \mu, \sigma)$ = probability of the statement "$\mu \epsilon f(v)$" being false, and $\beta(f \mid \mu, \sigma)$ = expected Lebesgue measure of $f(v)$. We consider the problem of constructing $f(v)$ so as to make both α and β "as small as possible." One of the results obtained is as follows.

Given p, $0 < p < 1$, let $f^0_{\lambda;\zeta(p)}(v) = \{\mu: \sum_1^k n_i[(\bar{x}_i - \mu_i)/l_i]^2 < \zeta(p) \cdot \sum_1^k n_i[s_i/l_i]^2\}$, where $\bar{x}_i = n_i^{-1} \sum_1^{n_i} x_{ij}$, $s_i^2 = n_i^{-1} \sum_1^{n_i} (x_{ij} - \bar{x}_i)^2$, $\lambda = (l_1, l_2, \cdots, l_k)$, the l_i's being given positive constants, and $\zeta(p)$ being determined by $P(\chi_k^2 > \zeta(p) \cdot \chi_{N-k}^2) = p$, where χ_k^2, χ_{N-k}^2 are independent chi-square variables with k, $N - k$ degrees of freedom ($k < N = \sum_1^k n_i$). Then (a) obviously $\alpha(f^0_{\lambda;\zeta(p)} \mid \mu, c\lambda) = p$ for all μ and all c, $0 < c < \infty$, and (b) if $f(v)$ is any other function such that $\alpha(f \mid \mu, c\lambda) \leq p$ for all μ and all c, either (i) $f(v)$ and $f^0_{\lambda;\zeta(p)}(v)$ differ by a set of measure zero for almost every v, or (ii) $\sup_{\mu \epsilon R} \{\beta(f \mid \mu, c\lambda)\} > \sup_{\mu \epsilon R} \{\beta(f^0_{\lambda;\zeta(p)} \mid \mu, c\lambda)\}$ for every c.

122 KAWADA, YUKIYOSI. (Tokyo University of Literature and Science). **Independence of Quadratic Forms in Normally Correlated Variables.**

An extension is given of theorems of Craig, Hotelling and Matérn which includes the following theorem, proved by a new method: If two quadratic forms Q_1, Q_2 in normally and independently distributed variates with zero means and unit variances satisfy the four conditions $E(Q_1^i Q_2^j) = E(Q_1^i)E(Q_2^j)$, for $i, j = 1, 2$, then the product of the matrices of the two forms in either order is zero.

123 VORA, SHANTILAL AMIDAS. (Institute of Statistics, University of North Carolina). **Bounds on the Distribution of Chi-Square.**

Let

$$\chi^2 = \sum_{j=1}^{j=k} (v_j - np_j)^2/np_j , \qquad \chi'^2 = \sum_{j=1}^{j=k} (v_j + \tfrac{1}{2} - Np_j)^2/Np_j$$

where

$$v_i \geq 0, \quad \sum_1^k v_i = n, \quad p_i > 0, \quad \sum_1^k p_i = 1 \quad \text{and } N = n + k/2.$$

Bounds on the multinomial probability T in terms of χ'^2 are obtained. A triangular transformation of

$$x_i = (v_i + \tfrac{1}{2} - Np_i)/\{Np_i(1 - p_i)\}^{1/2},$$

$(i = 1, \cdots, k - 1)$ to y_i is applied so that

$$d\chi'^2 = \sum_{i=1}^{i=k-1} y_i^2 ,$$

where d is determined later by equating the coefficients of χ'^2. Certain rectangles $r(v)$ with $(y_1 , \cdots , y_{k-1})$ as a mid-point are non-overlapping and cover the entire space R_{k-1} for $v_i = 0, \pm 1, \pm 2, \cdots$. If $\chi'^2 \leq c$, then bounds on T in terms of the integral of the $(k - 1)$ dimensional normal frequency function over the rectangle $r(v)$ are obtained. Prob. $\{\chi'^2 \leq c\}$ is the sum of T over $\chi'^2 \leq c$, so the integral over the sum of rectangles whose mid-points lie within the hypersphere $\chi'^2 \leq c$ is considered. Two hyperspheres, one which contains the sum of those rectangles, and one which is contained in it are used for the bounds, giving

$$\lambda_2 \cdot F_{k-1}(c_2) \leq \text{Prob.}\{\chi'^2 \leq c\} \leq \lambda_1 \cdot F_{k-1}(c_1),$$

where $F_{k-1}(x)$ is a chi-square distribution function with $(k - 1)$ degrees of freedom and $\lambda_1 , \lambda_2 , c_1 , c_2$ are functions of c, n, k and $p_1 , \cdots , p_k$. As $n \to \infty$, both bounds tend to $F_{k-1}(c)$. Bounds of the same form are obtained for Prob. $\{\chi^2 \leq C\}$. Closer bounds for Prob. $\{\chi^2 \leq C\}$ are given in terms of a non-central chi-square distribution.

124 HENDERSON, C. R. (Cornell University). **Estimation of Genetic Parameters.**

Many applications of genetics and statistics to the improvement of plants and animals deal with experimental data for which the underlying model is assumed to be

$$y_\alpha = \sum_{i=1}^{p} b_i x_{i\alpha} + \sum_{i=1}^{q} u_i z_{i\alpha} + e$$

where b_i are unknown fixed parameters, $x_{i\alpha}$ and $z_{i\alpha}$ are observable

parameters, the u_i are a random sample from a multivariate normal distribution with means zero and covariance matrix $|| \sigma_{ij} ||$, and the e_α are normally and independently distributed with means zero and variances σ_α^2. If $\sigma_{ij} = 0$ when $i \neq j$ and if $\sigma_\alpha^2 = \sigma_e^2$, the model is the one usually assumed when components of variance are estimated.

Three different estimation problems are involved, (1) estimation of b_i under the assumptions of the model, (2) estimation of u_i and (3) estimation of σ_{ij}. The first two problems are not solved satisfactorily by the least squares procedure in which the u_i are regarded as fixed, but the maximum likelihood solution does lead to a satisfactory estimation procedure.

Assuming that the σ_{ij} and σ_α^2 are known, the joint maximum likelihood estimates of b_i and u_i are the solution to the set of linear equations,

$$\sum_{i=1}^{p} b_i \Big(\sum_\alpha x_{h\alpha} x_{i\alpha} / \sigma_\alpha^2 \Big) + \sum_{i=1}^{q} u_i \Big(\sum_\alpha x_{h\alpha} z_{i\alpha} / \sigma_\alpha^2 \Big) = \sum x_{h\alpha} y_\alpha / \sigma_\alpha^2 \qquad h = 1, \cdots, p$$

$$\sum_{i=1}^{p} b_i \Big(\sum_\alpha x_{i\alpha} z_{h\alpha} / \sigma_\alpha^2 \Big) + \sum_{i=1}^{q} u_i (\sigma^{ih} + \sum_\alpha z_{i\alpha} z_{h\alpha} / \sigma_\alpha^2) = \sum_\alpha z_{h\alpha} y_\alpha / \sigma_\alpha^2 \qquad h = 1, \cdots, q$$

Some important applications of this estimation procedure to genetic studies are described and certain computational short-cuts are suggested.

The problem of estimating σ_{ij} has not been solved satisfactorily although under certain quite general assumptions the equations for the joint estimation of b_i, u_i, σ_{ij}, and σ_α^2 can easily be written. The solution to the equations, however, is too difficult to make the procedure practical. Nevertheless unbiased estimates of σ_{ij} can be obtained by equating to their expected values the differences between certain reductions in sums of squares computed by least squares and solving for the σ_{ij}. In general, the expectation of the reduction due to

$$b_1, \cdots, b_p, u_1, \cdots, u_k (k \leq q) \text{ is } \sum_{gh} \sum d^{gh} E(Y_g Y_h),$$

where d^{gh} are the elements of the matrix which is the inverse of the $(p + k)^2$ matrix of coefficients and the Y_g are the right members of the least squares equations.

125 COHEN, A. C., JR. (University of Georgia). **Estimating the Mean and Standard Deviation of Normal Populations from Double Truncated Samples.**

The method of maximum likelihood is employed to obtain estimates of the mean and standard deviation of a normally distributed population from double truncated random samples. Two cases are considered. In the first, the number of missing variates is assumed to be unknown. In the second, the number of missing (unmeasured) variates in each tail is known. Variances for the estimates involved in each case are obtained from the maximum likelihood information matrices. A numerical example is given to illustrate the practical application of the estimating equations obtained for each of the two cases considered.

126 KALLIANPUR, GOPINATH. (Institute of Statistics, University of North Carolina). **Minimax Estimates of Location and Scale Parameters.**

If the joint frequency function of the random variables $X_1 , \cdots X_N$ contains only a scale parameter and is of the form

$$\frac{1}{\alpha^N} p\left(\frac{x_1}{\alpha} , \cdots , \frac{x_N}{\alpha}\right),$$

then under mild restrictions the following theorem is proved.

Theorem 1. If the loss function is of the form $W[(\alpha - \bar{\alpha})/\alpha]$, the best or minimax estimate $\bar{\alpha}_0(x)$ minimizes the integral

$$\int_0^\infty W\left(\frac{\alpha - \bar{\alpha}}{\alpha}\right) \frac{1}{\alpha^N} p\left(\frac{x_1}{\alpha} , \cdots , \frac{x_N}{\alpha}\right) d\alpha$$

and further

$$\bar{\alpha}_0(\mu x_1 , \cdots , \mu x_N) = \mu \bar{\alpha}_0(x_1 , \cdots , x_N) \quad ; \quad \mu > 0.$$

When both location and scale parameters are present and the joint frequency function is of the form

$$\frac{1}{\alpha^N} p\left(\frac{x_1 - \theta}{\alpha} , \cdots , \frac{x_N - \theta}{\alpha}\right),$$

then (under conditions similar to those in Th. 1) one of the results obtained is

Theorem 2. If the loss function is of the form $W[(\theta - \bar{\theta})/\alpha]$, the best estimate $\theta_0(x)$ of θ minimizes the integral

$$\int_{-\infty}^\infty \int_0^\infty W\left(\frac{\theta - \bar{\theta}}{\alpha}\right) \frac{1}{\alpha^N} p\left(\frac{x_1 - \theta}{\alpha} , \cdots , \frac{x_N - \theta}{\alpha}\right) d\theta \, d\alpha$$

and

$$\bar{\theta}_0\left(\frac{x_1+\lambda}{\mu}, \cdots, \frac{x_N+\lambda}{\mu}\right) = \frac{\bar{\theta}_0(x_1, \cdots, x_N)+\lambda}{\mu}; \qquad \mu > 0.$$

These theorems have been applied to derive minimax estimates in the case of standard distributions. Finally, the problem of estimating the difference between the location parameters of two populations is briefly considered. The results obtained in this paper are a continuation of the line of approach suggested in Theorem 5 of Wald's "Contributions to the Theory of Statistical Estimation and Testing Hypotheses." *Ann. Math. Stat.*, Vol. 10, 1939.

The present work was carried out under ONR contract.

ROY, S. N. (Institute of Statistics, University of North Carolina).

127 On Some Features of the Neyman-Pearson and the Wald Theories of Statistical Inference, Their Interrelations and Their Bearing on Some Usual Problems of Statistical Inference.

With two alternative hypotheses H_1 and H_2 it is shown that (i) the most powerful test of H_1 with respect to H_2 is automatically an unbiassed test in the sense that its power is never less than (and usually greater than) the level of significance α and (ii) there is also a least powerful test with its power not greater (usually less) than α. This means that all tests have powers lying in between, which gives a complete picture of the possible family of tests and provides a basis for defining efficiency of tests.

With the first kind of error α is tied up a minimum second kind of error β (complementary to the maximum power P), and the level at which α is fixed depends upon some compromise between α and β. This intuitive approach is formalised by the introduction of loss functions related to, and apriori probability weights for H_1 and H_2, thus leading to the first stage in the Wald treatment of dichotomy with two solutions in the observation space corresponding respectively to minimum and maximum total risks. This is immediately generalised to the first stage in the Wald treatment of multichotomy with minimum and maximum total risk solutions. An important special case is discussed in which all the possible alternatives to a particular hypothesis are, by our test procedure, indistinguishable among themselves, thus effectively forming only one alternative to the hypothesis, which means a degenerate multichotomy. The bearing of this on most powerful tests on an average under the Neyman-Pearson theory is also discussed.

The problem of testing of composite hypothesis which is usually treated in terms of the Neyman-Pearson theory is posed and treated in terms of the (first stage) Wald theory and an indication is given of ho

these notions could be applied to the usual problems of univariate and multivariate analysis.

128 DAVIS, R. C. (U. S. Naval Ordnance Test Station, Inyokern, Calif.). **Note on Uniformly Best Unbiased Estimates.**

For the estimation in an absolutely continuous probability distribution of an unknown parameter which does not possess a sufficient statistic, it is shown that no unbiased estimate for the unknown parameter exists which attains minimum variance uniformly over a parameter set of arbitrary nature. This result demonstrates the impossibility of obtaining a generalized sufficient statistic first proposed by Bhattacharyya. Although not used in this note it is surmised that Barankin's powerful results on locally best unbiased estimates can be applied to yield further results in this direction.

129 ROBBINS, HERBERT E. (University of North Carolina). **Competetive Estimation.**

Let θ be a vector random variable with distribution function $G(\theta)$ and let x be a vector random variable whose frequency function $f(x;\ \theta)$ depends on θ. Two statisticians, A and B, are required to estimate θ from the value of x. If A's estimate is closer to θ he wins one dollar from B and *vice versa*; in case of a tie no money changes hands. It is shown that A should estimate θ by the function $a(x)$ = median of posterior distribution of θ given x; his expected gain will then be ≥ 0 whatever estimate B may use. If $G(\theta)$ is not known to A he should estimate it from the series of values of θ which have been observed in previous trials. If these are not known, A should estimate $G(\theta)$ from the values of x which have previously occurred; how this may be done is discussed elsewhere (see following Abstract).

From the point of view of the theory of games, when $G(\theta)$ is unknown we have a game in which the "rules" are unknown and must be successively estimated from past experience. Other examples arise whenever a game involves random devices whose probability distributions are not known to the players but must be inferred by statistical methods, in general from secondary variables which contain only part of the total information. The rôle of statistical inference in such "long term" games is fundamental.

130 CHAND, UTTAM. (Department of Mathematics, Boston University). **The Effect of an Unknown 'Location Disturbance' on "Student's" t Based on a Linear Regression Model.**

Consider $y_1, \cdots y_{N_1}, y_{N_1} + 1, \cdots y_N$, a set of observations ordered in time. If the y's are normally and independently distributed according to $N(\alpha + \beta(t - \bar{t}), \sigma^2)$ and we want to find out if the y's have changed with time, we usually employ a "Student" t type of statistic with $N - 2$ degrees of freedom. If, as a consequence of the impact of a certain unknown political or economic change in the past on the y's, the y's actually constitute two independent, normal samples $y_1, \cdots y_{N_1}, y_{N_1} + 1, \cdots y_N$ distributed according to $N(m_1, \sigma^2), N(m_2, \sigma^2)$ respectively, a two-sample "Student" t also based on $N - 2$ degrees of freedom would be the appropriate statistic to use for the hypothesis $m_1 = m_2$. If, in fact, the latter situation describes the correct state of affairs, and the statistician employs the "Student" t based on the regression model, he commits an error. The present paper investigates the nature of such an error in the light of the point of impact as determined by the magnitude of N_1 and the intensity of the impact as determined by the standardized

$$\text{'distance'} = \frac{m_2 - m_1}{\sigma \sqrt{\frac{1}{N_1} + \frac{1}{N - N_1}}}$$

of this extraneous 'shock' on the ordered set of observations y.

131 BRADLEY, RALPH A. (The University of North Carolina and McGill University). **Corrections for Non-Normality for the Two-Sample t and the F Distributions Valid for High Significance Levels.**

The effects of non-normality of the parent population on common tests of significance have long been of concern in the application of statistical methods to experimental data. In this paper, the two-sample t statistic is expressed as a simple multiple of the cotangent of an angle between two lines in a space of dimensionality one less than the total of the sample sizes; the F statistic for k samples is expressed as a multiple of the cotangent of an angle between a line and a plane of $(k - 1)$ dimensions in a space, again, of dimensionality one less than the total of the sample sizes. The geometrical formulation is such as to suggest approximations to the distributions of these statistics valid for large values of the statistics, and these approximations are obtained. The approximations are shown to be exact in the special cases where the parent population is normal, and a method of evaluation of correction factors is given for a wide class of parent populations. The approximation procedures are valid for the distributions under both null and non-null hypotheses.

HANNAN, JAMES F. (Institute of Statistics, University of North Carolina). **132 Some Tests Based on the Empirical Distribution Function (Preliminary Report).**

Let $\underline{X} = (X_1, X_2, \cdots, X_n)$ be an independent sample of n where X_i has the continuous c.d.f. $F(x)$. Let $S_n(x)$ be the empirical distribution function. Acceptance regions of the type $\{\underline{X}: S_n(x) \leq \phi(x) \text{ for all } x\}$ are considered for different specifications of ϕ and their probabilities evaluated. The method of evaluation consists in identifying the regions with regions defined in terms of the order statistics of a sample of n from the uniform distribution on the interval (0, 1). The result obtained for $\phi(x) = F(x) + c/n$ $0 \leq c$ integral $\leq n$ is used to provide a direct proof of the Kolmogoroff result

$$\lim_{n=\infty} P[n^{1/2} \sup_x (S_n(x) - F(x)) \leq z] = 1 - e^{-2z^2},$$

while that obtained for $\phi(x) = F(x) + t$, $0 < t < 1$ gives the exact c.d.f. of the statistic $\sup_x (S_n(x) - F(x))$.

WALSH, DR. JOHN E. (The Rand Corporation). **133 On a Generalization of the Behrens-Fisher Problem.**

Let $m + n$ independent observations be available where it is only known that a specified m of them are from continuous symmetrical populations with common median μ while the remaining n are from continuous symmetrical populations with common median ν. This is the generalization of the Behrens-Fisher problem investigated; some tests and confidence intervals for $\mu - \nu$ which are valid for the generalized situation are presented. For definiteness, suppose that $n \leq m$. The procedure used is to subdivide the m observations (common median μ) into n groups of nearly equal size and form the mean of the observations for each group. Pair the n means with remaining n observations and subtract the value of each observation from the value of the mean with which it is paired. The resulting n values represent independent observations from populations with common median $\mu - \nu$. Tests and confidence intervals for $\mu - \nu$ are obtained by applying the results of "Applications of Some Significance Tests for the Median Which are Valid Under Very General Conditions" (*Amer. Stat. Assoc. Jour.*, Vol. 44, 1949, pp. 342–55) to these n values. To measure the "information" lost by using the generalized tests when one actually has two independent samples from normal populations, power efficiencies are computed with respect to: (a) Scheffe's "best" t-test solution and (b) Most powerful solution when ratio of variances is known. Case (a) yields an upper bound while case (b) furnishes a lower bound for the actual efficiency.

134 SHRIKHANDE, S. S. (Institute of Statistics, University of North Carolina). **Construction of Partially Balanced Designs with Two Accuracies.**

Various methods of construction of Partially Balanced Designs first introduced by Bose and Nair (*Sankhyā*, 4 (1939), pp. 337–373) have been considered. Two of the methods given are generalisations of a Difference Theorem given by them. Another method is the inversion of an unreduced Balanced Incomplete Block Design with $k = 2$. Use has also been made of the existing Balanced Incomplete Block Design in another direction. A number of designs can also be obtained by methods of Finite Geometries and especially by omitting a number of treatments and certain blocks from the complete Lattice Designs. Use of curves and surfaces in Finite Geometries and the use of multifactorial designs given by Plackett and Burman (*Biometrika*, 33 (1946), pp. 305–325) are also indicated.

135 SHRIKHANDE, S. S. (Institute of Statistics, University of North Carolina). **Designs for Two-Way Elimination of Heterogeneity.**

Use has been made of the existing Balanced and some Partially Balanced Designs for two-way elimination of heterogeneity with at most two accuracies. Particular cases of these designs were given by Youden (Contributions from Boyce Thompson Institute, IX (1937), pp. 317–326) and Bose and Kishen (Science and Culture (1939), pp. 136–137). The method depends upon interchanging the positions of various treatments in the different columns (blocks), if necessary, so as to satisfy certain conditions.

136 SHRIKHANDE, S. S. (Institute of Statistics, University of North Carolina). **Designs for Animal Feeding Experiments.**

In animal-feeding experiments change-over designs are generally preferable to continuous feeding experiments. In change-over designs both the direct and carry-over treatment effects are important. Use of Balanced and Partially Balanced Incomplete Block Designs toward this end has been considered.

137 SANDELIUS, D. MARTIN. (Statistiska Institutionen, Uppsala Universitet). **A Truncated Sequential Procedure for Interval Estimation with Applications to the Poisson and Negative Binomial Distributions (Preliminary Report).**

Let x, y_1 , y_2 , $\cdots$ be a sequence of random variables defined in $(0, \infty)$, and let n be the smallest integer satisfying $\sum_{i=1}^{n+1} y_i > tx$, where

$t > 0$ is a non-random quantity. Define u_k either as $\sum_{i=1}^{k} y_i/x$ or as the smallest integer exceeding $\sum_{i=1}^{k} y_i/x$, $k = 1, 2, \cdots$. Given the distribution function $F(x, \theta)$ of x and, for any t, the conditional distribution of n with respect to x, the distribution of u_k is obtained. The problem is to determine a confidence interval for θ with confidence coefficient $1 - \alpha$ on the basis of either an observation on u_k, if $u_k \leq t$, or an observation on n, if $n \leq k - 1$. The following procedure is proposed: If $u_k \leq t$, choose θ_{10} and θ_{11} according to a rule satisfying Prob $(\theta_{10} \leq \theta \leq \theta_{11} \mid u_k \leq t) \geq 1 - \alpha$. If $n \leq k - 1$, choose θ_{20} and θ_{21} such that Prob $(\theta_{20} \leq \theta \leq \theta_{21} \mid n \leq k - 1) \geq 1 - \alpha$. For continuous u_k the following cases are discussed: A) $x = \theta$ with probability 1, and n has, for any t, a Poisson distribution with mean $t\theta$, B) x has a Gamma distribution with mean θ, and the conditional distribution of n with respect to x is, for any t, a Poisson distribution. Both cases may, for instance, be applied to bacterial counting.

138 ROBBINS, HERBERT E. (University of North Carolina). **A Generalization of the Method of Maximum Likelihood: Estimating a Mixing Distribution (Preliminary Report).**

Let θ be a vector random variable with distribution function $G(\theta)$ belonging to some class G, let x be a vector random variable whose frequency function $f(x; \theta)$ depends on θ, and let $g^*(x) = \int f(x; \theta) dG(\theta)$ be the resulting frequency function of x. From a sample $x_1, x_2, \cdots$ it is required to estimate $G(\theta)$. The generalized method of maximum likelihood consists in using the estimates $G_n(\theta; x_1, \cdots, x_n)$ in G for which $\prod_{i=1}^{n} g^*(x_i)$ is a maximum. Under certain restrictions this method is consistent as $n \to \infty$. More generally, if the distribution function of x is of the form $G^*(x) = \int F(x; \theta) dG(\theta)$ and if this integral equation with kernel $F(x; \theta)$ defines a one-to-one continuous correspondence between $G(\theta)$ and $G^*(x)$, then $G(\theta)$ can be estimated by using the sample distribution function of $x_1, \cdots, x_n$ to replace $G^*(x)$ and solving for the corresponding $G(\theta)$. We can also apply the method of minimizing in G an appropriate measure of the deviation between the sample distribution function of $x_1, \cdots, x_n$ and $G^*(x)$.

Any consistent method of estimating the mixing distribution $G(\theta)$ from the sequence $x_1, x_2, \cdots$ yields a solution of parametric statistical decision problems in the following manner: from *past* values $x_1, \cdots, x_{n-1}$ we estimate $G(\theta)$, and then use the corresponding Bayes solution of the decision problem to reach our decision for x_n, even though the value θ_n which produced x_n is different from those which produced $x_1, \cdots, x_{n-1}$. In certain cases of long-term experimentation this approach seems more reasonable than the minimax method which decides on the course of action appropriate to θ_n on the basis of x_n only, and ignores the information about the prior distribution of θ which is contained in $x_1, \cdots, x_{n-1}$.

THE BIOMETRIC SOCIETY

Officers for 1950. According to our constitution, the general officers are elected by Council each year, with an obligatory change in President every two years. In accord with these provisions the Council has elected Arthur Linder as President, J. W. Hopkins as Treasurer and C. I. Bliss as Secretary for 1950. Our new President is professor of mathematical statistics in the University of Geneva and in the Swiss Federal Institute of Technology at Zurich. Much of the success of the Second International Biometric Conference last summer in Geneva was due to the skill and tact with which he handled the many arrangements for the Conference.

The following Council members were elected for the period 1950–52 inclusive by mail ballot of the members of the Society: M. H. Belz, G. Darmois, R. A. Fisher, P. V. Sukhatme, O. Tedin and E. B. Wilson. J. W. Trevan was named to complete President Linder's unexpired term as member of Council. Dr. Jane A. Russell and J. R. Wittenborn of Yale University served as tellers.

Meetings in Stockholm. The International Union of Biological Sciences will meet this coming summer in Stockholm, Sweden on July 7–11. Since the Biometric Society provides the secretariat of its Section on Biometry, we will be represented on the Executive Committee of the IUBS by President Linder and one other member of the Society. An International Congress of Botany on July 12–20, also in Stockholm, will follow the IUBS meetings. A biometrical program is planned in connection with these two meetings which we hope will aid in the development of a Scandinavian Region in the Society.

BIOMETRICS. Since its formation, the Biometric Society has used BIOMETRICS as its journal, although ownership resided in the American Statistical Association and its Biometrics Section. As an international organization, the Society has felt the need of having its own journal. Because of the contributions to BIOMETRICS by the Society, both in content and subscriptions, the Council proposed to the American Statistical Association that the journal be transferred to the Society with appropriate safeguards for the interests of the members of the ASA who subscribe to BIOMETRICS but are not members of the Society. Transfer of BIOMETRICS to the Society was approved successively by the ASA Biometrics Section, Board of Directors and Council at meetings in December 1949. President-elect Lowell J. Reed and the Chairman of the Biometrics Section, H. F. Dorn, were named as representatives of the

ASA to arrange the transfer. The Council of the Biometric Society has asked J. W. Hopkins, C. I. Bliss and Gertrude M. Cox to represent the Society in arranging the transfer. When negotiations have been completed, details of the transfer will be published in BIOMETRICS.

Proceedings of a biometrical-entomological clinic. A biometrical clinic for entomological problems was held in New York on December 13, 1948, under the sponsorship of the Association of Economic Entomologists and the Eastern North American Region of the Society. The proceedings were recorded electronically, transcribed and edited. The Council of the Society approved their being published through the Secretary's office if a sufficient number of prepublication orders were received to cover the estimated expense. In response to notices sent to the members of both organizations, 98 orders were received from Society members and 489 orders from the economic entomologists. An edition of 700 copies was printed in February as a multilithed 64-page bulletin. As long as the supply lasts, copies can be obtained from the office of the Secretary at 50¢ for members of both sponsoring organizations and 75¢ for others.

ENAR. The Eastern North American Region held its annual meeting in New York City on December 28–30 jointly with the Biometrics Section of the American Statistical Association and the American Association for the Advancement of Science. At the Regional business meeting on December 30, the following officers were named for 1950: Vice-President, Joseph Berkson; Secretary-Treasurer, Walter T. Federer; Members of the Regional Committee from 1950 to 1952, Lila F. Knudsen and W. J. Youden. The scientific program consisted of three sessions. The first under the chairmanship of H. W. Norton concerned the use of rationally developed equations in biology, with papers by S. E. Luria and by E. V. Newman and Margaret Merrell. The second on long-time follow-up in morbidity studies was chairmaned by J. W. Fertig and included papers by P. M. Densen, by T. E. Harris, Paul Meyer and J. W. Tukey and by H. F. Dorn. The third session with Frederick Mosteller as chairman consisted of contributed papers by Joseph Berkson, by Jane Worcester and S. S. Stevenson, by W. J. Youden and by S. W. Greenhouse and Nathan Mantel.

The Biometric Society, E.N.A.R. and the Institute of Mathematical Statistics met jointly at Chapel Hill, North Carolina, March 17–18, 1950. All sessions were joint meetings of the two organizations. Forty members of the Biometric Society were in attendance.

Two papers due particular comment are the invited addresses on "Mathematical Models in Biology" by Dr. James A. Rafferty and on "Estimation of Genetic Parameters" by Professor C. R. Henderson. Dr. Rafferty stressed that much more cooperation among biologists,

mathematicians and statisticians is needed, if biological research is to go forward at the pace required by modern living. Professor Henderson outlined the estimation methods currently used by animal breeders, pointed out their difficulties and deficiencies, and asked the aid of mathematical statisticians in improving these methods.

Highlights of the meeting included a dinner on Friday evening at the Carolina Inn. Professor W. G. Cochran was toastmaster, and welcome was bid the two societies by Chancellor Robert B. House of the University of North Carolina. Professor Gertrude M. Cox responded for the Biometric Society and Professor David F. Votaw for the Institute of Mathematical Statistics.

At the close of the meetings, Saturday afternoon, the attendees were guests for tea at the home of Professor and Mrs. Harold Hotelling. Many visited the Morehead Building and enjoyed the spectacular show "Easter Awakening" given at the Planetarium.

Professor Harold Hotelling was in charge of program for the Institute of Mathematical Statistics, and Professor H. L. Lucas in charge of program for the Biometric Society. Professor H. E. Robbins was in charge of arrangements and accommodations, and was assisted by Professor George E. Nicholson. These two should be commended for doing an heroic job.

Two other regional meetings were held in April. The first was a joint session with the Society of Pharmacology and Experimental Therapeutics at Atlantic City on April 19. Practical problems submitted by pharmacologists were discussed informally by a panel consisting of C. I. Bliss, W. G. Cochran, Frederick Mosteller and W. J. Youden, with E. G. deBeer as chairman. The second was a joint meeting with the American Mathematical Society at Oak Ridge, Tennessee on April 21, where the program included three papers on topics in bio-mathematics by J. Z. Hearon, by A. Rapoport and by C. W. Sheppard.

Indian Region. The Indian Region held its annual meeting in Poona on January 5, 1950. The following were elected to serve on the governing Council for the year 1950: Vice-President, P. C. Mahalanobis; Secretary, C. Radhakrishna Rao; Treasurer, Anukul Chandra Das; Members, V. M. Dandekar, K. Kishen, K. R. Nair, U. S. Nair, V. G. Panse, P. B. Patnaik, B. Ramamurthy, F. N. Roy and P. V. Sukhatme. A scientific program included papers by V. M. Dandekar, by Basudeb Banerjee and Anukul Chandra Das, by S. Raja Rao, by S. Janardan Poti and by C. Radhakrishna Rao. The meeting voted to invite the Biometric Society to hold an international session in India in the fall of 1951.

Région Française. Région Française held its annual meeting on February 28 at the Laboratory of Zoology of the Faculty of Sciences in Paris. The three ordinary members of the Regional Council were re-

elected, M. Lamotte, Mlle. Colette Rothschild and M. P. Schutzenberger. The scientific sessions included three papers on problems of correlation, two of them by P. Rey and the third by M. P. Schutzenberger. A second session of the Region late in April considered the probability of an all-or-none response as a function of a dose or other parameter.

British Region. At the annual meeting of the British Region on March 14, the following officers were named: Vice-President, R. A. Fisher; Treasurer, A. R. G. Owen; Secretary, D. J. Finney; Regional Committee Members for 1950–1952, W. L. M. Perry and C. B. Williams. Abstracts of the papers read at the meeting appear in this issue of BIOMETRICS.

The Interests of Members of the Biometric Society. We are indebted to Professor John W. Tukey of Princeton University for the following analysis of the widespread interests of our members.

The directory published last summer lists the interests of 93% of the 900 who were members at that time. No interest was available for 62 (30AR, 11BR, 7IR, 7 unattached, 3RF, 3ENAR, 3WNAR). A rough examination of the addresses and titles of these 62 suggests that their probable interests resemble those of the other 838, and so all members below are based on 838. Many members stated two, three or even four interests. They have been prorated. Thus, Wilcoxon, who gave "insecticides, fungicides, herbicides, biometry", was counted as 2/4 under "entomology, mycology, parasitology, nematology", 1/4 under "agriculture", and 1/4 under "biometry". There seemed to be no easily applied rule that would be fairer than this. (When information is gathered for another directory, it may be possible for each member to give better weighted information.)

According to its letterhead, the Biometric Society is "An international society devoted to the mathematical and statistical aspects of biology". We should hope then, that most interests would be biological, and second would come mathematics and statistics. This is indeed the case, for we find

Biological sciences	53%	(441.4)
Mathematics and statistics	22%	(184)
Bridge fields of biology	17%	(139.8)
Other areas	8%	(72.8)

Not only this broad breakdown, but also the more detailed distribution is of interest to members, officers and editors. Not all will agree with this classification, and so considerable detail is useful, since it permits ready rearrangement.

In the detailed breakdowns, the percentages are based on the totals of the broad classifications.

BIOLOGICAL SCIENCES (441.4)

22% Genetics (96 as follows: unspecified 42.3; plants 17.5; human 12; animals 10.5; poultry 4; microorganisms 1; quantitative 8.7).
14% Bioassay and pharmacology (62.6 as follows: bioassay and antibiotics 35.8; pharmacology, therapeutics and toxicology 26.8).
12% Medicine and public health (55 as follows: medicine, clinical research, etc. 14.3; surgery, pathology and specialties 16.7; hygiene and industrial health 8.5; public health 15.5).
12% Physiology and nutrition (51.8 as follows: physiology 32; nutrition 19.8).
11% Agriculture, forestry, fisheries (49.3 as follows: agriculture 22.8; forestry, horticulture and pomology 16; fish and aquatic biology 10.5).
8% Human specialties (34.3 as follows: psychology 22.8; human biology, anthropology and biotypology 11.5).
15% Other specialties (66.7 as follows: evolution, population genetics and ecology 17.5; entomology, mycology, parasitology and nematology 17.4; microbiology, bacteriology and virology 12.3; serology, hematology and immunology 7.3; cytology, embryology, anatomy and histology 6.2; herpetology and bird biology 6).
6% General biology (25.7 as follows: unspecified and applied 6.2; botany, zoology and systematics 10.3, quantitative, theoretical and mathematical 9.2).

MATHEMATICS AND STATISTICS (184)

92% Statistics (170 as follows: mathematical statistics and probability 71.8; unspecified and applied 69.5; experimental design, inference and scientific method 19.7; sampling 9).
8% Mathematics (14).

BRIDGE FIELDS OF BIOLOGY (139.8)

56% Statistical (78.5 as follows: vital statistics, demography and actuarial 24.3; medical statistics 19.2; public health statistics 16.7; biostatistics 12.8; agricultural statistics and agricultural economics 5.5).
24% Biometrics (33).
20% Physical science (28.3 as follows: biochemistry 16; biophysics and radiation biology 12.3).

OTHER AREAS (72.8)

30% Engineering (21.3).
28% Economics (20.5).
16% Chemistry (11.5).
11% Education (8.2).
7% Physics and geophysics (5.3).
8% Other sciences and technologies (6).

In interpreting these tables, we must remember that a figure like "37" may represent partial interests of 50, 60 or even 70 members.

RECENT APPLICATIONS OF BIOMETRICAL METHODS IN GENETICS

(1) EXPERIMENTAL TECHNIQUES IN PLANT IMPROVEMENT

F. Yates
Rothamsted Experimental Station
Harpenden, Herts, England

I was only asked to contribute to this discussion a few weeks before the conference, and pressure of other work has prevented me from tackling certain investigations on which I hoped I might make some progress. The present paper, therefore, is more a statement of problems which require solution, than a presentation of any definite conclusions. The problems I want to consider are those which arise when planning a testing scheme which will be effective in testing the large number of new lines and varieties which are produced in the course of any programme of plant improvement. In the main I have followed the lines of a paper I gave to a Congress of Plant Breeders which was held at the John Innes Horticultural Institution last November. To those who were present at that meeting I must offer my apologies. I am, however, encouraged to repeat what I said there, because only a very brief summary of the meeting has been published (Lewis, 1949), and because the present audience is largely different.

The reaction of the meeting itself was that the subject was of considerable importance and should be more fully discussed at the next Congress. Subsequent events showed, however, that the need for further discussion was not felt by all plant breeders, for another Congress was held this spring to which such troublesome characters as mathematical statisticians who might originate discussions of this type were not invited.

The problems which I posed at the earlier meeting therefore remain. They have not been further discussed, and I personally have not made any progress with their solution. As far as I know no one else has either. But ignoring a problem does not get rid of it. I will therefore propound the problems again, in the hope that some here may be stimulated to making a contribution towards their solution.

DESIGN OF EFFICIENT TESTING MECHANISMS

At all stages of the selection process we shall require to compare the various new lines and varieties that are produced in order to see which are sufficiently promising to be retained, either as a basis of further breeding work or as possible new varieties for commercial use. Quantitative characters such as yield must be tested in field trials, which may be regarded as the testing mechanism of plant breeding. As in all branches of science improvement of measuring instruments and testing mechanisms may be expected to result in progress in the science itself.

The design of field trials for comparison of different varieties is a branch of experimental design and is the only aspect of the subject under discussion which has received much consideration by mathematical statisticians. Development during the inter-war period has been considerable. The introduction of the principle of randomization by Professor R. A. Fisher was the first and fundamental step—replication and local control (arrangement in blocks or other systematic patterns) had of course long been in use, though in Great Britain at least there had been a retrogression to large plots without replication. Randomization laid the foundation of further advances by providing an unequivocal means of estimating the experimental errors. This had a number of important consequences, one of which was that it became possible to judge the relative efficiency of different types of design in an objective manner.

The second major development was the recognition of the importance of factorial design and the development of methods appropriate to elaborate factorial experiments—confounding, estimation of error from high-order interactions and fractional replication. The third major development, which was of particular interest to plant breeders, was the introduction of quasi-factorial and balanced incomplete block designs. These designs enable a large number of varieties (or treatments) to be compared in groups without the use of controls, thus permitting a greater degree of elimination of soil heterogeneity than would be possible with ordinary randomized blocks, while at the same time giving comparisons between pairs of varieties which are all of approximately equal accuracy, and computations which are reasonably simple.

I do not wish to discuss the various alternative incomplete block designs at length here, but I would like to make a few general points. In the first place the type of design which is most appropriate depends on the nature of the tests we desire to make. It has been suggested, for instance, that since incomplete block designs are *modern*, anyone using designs involving controls is somehow behind the times. This, of course, is not necessarily true. Under certain circumstances comparison

with standard varieties may be required; in such cases it may well be advantageous to use these as controls. Incomplete block designs are also unsuitable if a number of the strains to be tested are likely to fail completely, or if a number can be rejected by inspection without harvesting.

Secondly there is an upper limit to the gain in efficiency that can be obtained by elaboration of the experimental layout. This limit is set by the inherent variability of the individual plants due to factors which operate substantially at random from plant to plant, and to the residual component of fertility variation which occurs with even the smallest plots. Probably the least variable type of element for a given size and shape of plot is the Latin square doublet:

A B

B A

A comparison of the variability per plot of such doublets with the variability of designs for testing large numbers of varieties can be used to assess their inherent efficiency. But we must not expect to obtain efficiencies of 100 per cent, since there will always be some loss of efficiency with a large number of varieties owing to the number of different comparisons on which information is required. Thus if doublets of the above type are used to compare v varieties in balanced pairs the efficiency factor, if inter-doublet information is not utilized, is $v/2(v-1)$ or, if v is large, $\frac{1}{2}$. Other designs such as lattice designs in blocks and lattice squares may be expected to remove fertility differences somewhat less effectively than Latin-square doublets, and will consequently have a variance per plot which is somewhat greater than that for a doublet, but they will have higher efficiency factors. The efficiency factor of a set of 7×7 lattice squares (49 varieties) for example, is $\frac{3}{4}$. If the variance per plot is the same as for a doublet the inherent efficiency is 75 per cent plus whatever information is derivable from inter-row and inter-column comparisons. This latter information in its turn cannot exceed 25 per cent, and if there are marked fertility irregularities which are eliminated by the rows and columns, will be very much less.

I have not carried out any examination of the inherent efficiencies of lattice and lattice square designs, but I would hazard the guess that they will be found on the average to be not very far removed from 100 per cent. If this is so it is idle to look for more precise types of design.

There remains size and shape of plot. Up to a point the smaller the plots the greater the amount of information *per unit area*. Also up to a

point long narrow plots are more efficient than square ones of the same area. Limits are set in both cases, however, by the need for rejecting edge rows to avoid inter-plot competition; and even if edge rows do not have to be rejected inter-plant competition may result in the amount of information being maximum at some plot size well above the minimum of a single plant. If the greater part of the costs are proportional to the number of plots and not to total area maximum efficiency will be obtained with considerably larger plots.

One way of reducing the variability is to choose particularly uniform land for trials. This is good practice—indeed essential—as far as the central station is concerned, but the necessity of testing and to a certain extent selecting varieties under conditions similar to those in which they will be grown limits possibilities in this direction.

The other major advance of experimental design referred to above, namely factorial design, has not been nearly as fully exploited as it should have been in plant selection. Factorial design is, of course, not of direct use, since the different varieties or strains cannot constitute more than a single factor. Far more rapid agricultural progress can, however, be made by including other factors such as fertilizers, time of planting and so forth in varietal trials. This is of value in the plant selection work itself—thus, varieties of cereals are urgently required which will stand up to heavy nitrogenous manuring without lodging, and the respective merits of different varieties in this respect can only be directly tested by a factorial design involving both varieties and levels of nitrogenous manuring. It is also of value in that it leads to indirect economies by providing information on the other factors, thus rendering independent trials on these factors unnecessary. Factorial designs of the conventional type are likely to be of greatest use at the later stages of the testing programme. At the earlier stages, when large numbers of lines or varieties are involved, the other factors can well be applied to whole blocks, e.g. those of incomplete block designs.

Having evolved an efficient testing mechanism we must concern ourselves with the ways in which it can be applied in practice. It is this aspect of the matter which, it seems to me, has as yet been given very inadequate consideration, and it is here that we expect to see large increases in efficiency.

BALANCE BETWEEN THE DIFFERENT STAGES OF THE SELECTION PROCESS: NUMBER OF STAGES, PROPORTION OF VARIETIES OR LINES RETAINED AT EACH STAGE, ACCURACY AT EACH STAGE

The most efficient routine for plant breeding and selection will depend very much on the genetical situation, and consequently the optimal

methods for any given crop will have to be evolved by close cooperation between mathematical statisticians and geneticists engaged in the breeding of that crop. The maintenance of breeding stocks also presents somewhat different problems from the production of new commercial varieties. In the former the preservation of genetic potentialities is required, as well as progress in the desired direction. In a new commercial variety suitability to the particular needs of the moment is the governing factor. The essence of the situation in all crops, however, is that the production of new untested strains is a relatively simple job. It is the testing that involves the work. It may well, therefore, pay to produce a large number of lines, carrying out the initial tests with low accuracy and gradually refining the selection. This is what is in fact done intuitively by all plant breeders.

Even with a single stage of selection at each generation the greater the number of lines, the greater the chance that some particularly good line is included. On the other hand, comparisons between the individual lines will be less accurate when a larger number are included, owing to the smaller number of plots that can be devoted to each line, and owing to the increasing error per plot as the number of varieties is increased.

In 1939 at the 7th International Congress of Genetics, I pointed out (Yates, 1940) that neglecting the latter factor entirely, and assuming that the total number of plots is fixed, the average genetic advance due to the selection of the apparently best variety, instead of a random variety, where there is no retrogression in the absence of selection, will be

$$\frac{G}{G + \lambda n}\,\bar{x}_n\,,$$

where n is the number of varieties,

G is the genetic variance (distribution assumed normal),
λn is the experimental error variance, λ being independent of n,
$\bar{x}_n$ is the mean value of the greatest deviate of a sample of n from a normal population with unit standard deviation. (Tabulated in Fisher & Yates, Statistical Tables. 1938.)

With $\lambda = \frac{1}{10}G$ the optimum number of varieties will be 13, in which case the genetic variance will be somewhat less than the experimental error variance. With $\lambda = \frac{1}{100}G$ the optimum number will be somewhat greater than 50, and the genetic variance will be about twice the experimental error variance. In terms of the ordinary analysis of variance the variance ratios between varieties and error will average about 1.8 and about 3 respectively. The former is less than the value required to give significance at the 5 per cent point.

These simple considerations serve to emphasize the value of testing a large number of varieties with moderate accuracy instead of only a few with very high accuracy. In any series of trials involving only a few varieties which give varietal differences that are large compared with their standard errors the question should always be asked: would not the work have been improved if the same experimental resources had been devoted to the comparison of a larger number of varieties?

The above thoughts were prompted by the habit which had become common in certain quarters of judging the efficiency of a variety trial by the degree of significance obtained. I had hoped that others would continue the investigation, but so far as I know nothing further has been done.

The number of stages introduced into the testing process at each generation presents similar problems. If we have a thousand new lines, for example, we can, if we wish, test all lines simultaneously, or we can test them in a series of stages. Thus we might use three stages, retaining $\frac{1}{10}$ at each of the first two stages and ending with 10 lines between which reasonably accurate comparisons are available; or we might use two stages, retaining $\frac{1}{10}$ at the first stage and ending with 25 lines; or we might use two stages, retaining $\frac{1}{10}$ at the first stage and ending with 100 lines. Frankly, I have no idea of the relative merits of these various procedures, but I suspect that they may be very different.

Testing in stages is analogous to sequential analysis in sampling procedure. One disadvantage, which must not be forgotten, is that it lengthens the testing process. This may be serious in certain circumstances in agriculture, since each test normally occupies a season.

The possibility of testing the suitability of parental lines by testing their progenies in a number of different crosses also introduces further interesting statistical problems which are in some ways analogous to factorial design. If in a cross-fertilized plant, for instance, we test all reciprocal crosses between 20 parents (380 lines) we shall obtain very accurate measures of the average merit of a parent, even though the tests of the individual lines are of low precision. The analysis of this type of data has already been discussed (Yates, 1947). A further problem that still requires consideration is what weight to give to the average parental scores, and what to the scores of the individual lines, when arranging the lines in order of merit.

STANDARD VARIETIES

In the introduction of new varieties of an established agricultural crop we are usually interested in comparing them with established varieties. These established varieties can often well serve as controls in

the trials. Control varieties provide a standard for measuring the relative merits of groups of lines produced in different years. I would emphasize, however, that if control varieties are used, three or more should be used and not, as is often the case, a single variety. Any one variety may have peculiarities such as an abnormal reaction to certain meteorological conditions, particular susceptibility to certain diseases, etc. Again in vegetatively propagated plants such as potatoes a standard variety may become infected by a virus and gradually deteriorate for this reason. Consequently, standard varieties, when they are used, should be gradually supplanted by the superior new varieties, which in turn become the standards.

If the field trial testing scheme is not to become unmanageable varieties should not be permitted to enter the field trial stage until they have passed the necessary preliminary tests, but it is equally important that potential new varieties should be introduced into the field trials at an early stage. This is a defect of the present system of testing varieties in the United Kingdom, where the official tests of the National Institute of Agricultural Botany always tend to deal with varieties which have, in fact, been in current use by the more progressive farmers for a number of years on the recommendations of their seed merchants, recommendations which are frequently, I suspect, based on somewhat inadequate evidence or on trials in other countries from which the new varieties have been imported.

There must also be some orderly method of eliminating varieties from field trials, and, I trust, from general agricultural use, when they have got to the stage of being superseded by newer and better varieties.

QUALITY FACTORS

Testing for quality factors is of the utmost importance. It presents considerable practical difficulties, and has in consequence tended to be neglected. One trouble is that a fair bulk of produce is often required for quality testing. This may require the use of larger plots than would be the case if only yield was of importance. Certain quality factors such as flavour can only be tested by subjective means. This may have to be done by studying an order of preference, i.e. by ranking. C. I. Bliss has recently been doing some interesting work on these lines, using incomplete blocks.

ENVIRONMENTAL FACTORS

In order to test the suitability of different varieties to different soils and climatic conditions, tests must be carried out at different centres. Consequently we must organise a chain of experimental sta-

tions, or alternatively a series of trials on widely dispersed commercial farms. Agronomic factors such as amounts of fertilizers, sowing date etc. can be, and should be, dealt with in the variety trials themselves by the methods of factorial design mentioned above. For certain factors such as lodging, disease response, etc. it may be possible to develop special laboratory or greenhouse tests. Such tests are of the greatest value as they lead to sharpened and more speedy criteria of selection, and enable much more rapid progress to be made. They must, however, be confirmed by correlation with field trials.

CONDITIONS UNDER WHICH SELECTION IS MADE

The conditions under which the earlier stages of selection are made are, I suspect, sometimes at fault. To take a specific example, the stiffness and shortness of straw are very important properties in cereals. Yet, so far as I know, the first stage of selection of new lines is generally made from plants sown in close proximity and judged on their individual merits, particularly by apparent vigour. Thus, there is at this stage, as there is in a wild population, considerable selective advantage in height. This seems to be a situation in which selection under deliberately unnatural conditions at the early stages might well result in considerable advances, but I leave this to the geneticists.

REFERENCES

Fisher, R. A. and Yates, F. 1938. *Statistical tables for biological, agricultural and medical research.* Oliver & Boyd, Edinburgh. (3rd edition 1948).

Lewis, D. 1949. Problems and policy in plant breeding. *Nature, 163,* 51–53.

Yates, F. 1940. Modern experimental design and its function in plant selection. *Emp. J. Exp. Agric., 8,* 223–230.

Yates, F. 1947. The analysis of data from all possible reciprocal crosses between a set of parental lines. *Heredity, 1,* 287–301.

RECENT APPLICATIONS OF BIOMETRICAL METHODS IN GENETICS

(2) THE ANALYSIS OF SELECTION CURVES

Luigi L. Cavalli
Dept. of Genetics, The University, Cambridge

Quantitative data on spontaneous and artificial selection exist in the genetical literature, but no special attempts appear to have been made to produce statistical methods for their analysis. In this paper a method will be given for the analysis of a certain type of selection curves, namely those arising when two alleles in competition come to an equilibrium after a sufficient number of generations, the frequency of each allele at equilibrium being independent of the initial frequencies. The method can be easily extended to cover other types of selection processes.

The usual genetical hypothesis in this case is that the heterozygous genotype Aa, is at an advantage over both homozygotes AA, aa. The equilibrium will then depend on the relative survival values of the two homozygotes (Fisher, 1922, Haldane, 1926).

THE SELECTIVE PROCESS

Let p and $q = 1 - p$ be the gene frequencies of alleles A and a. Let the relative survival values of the three genotypes be

genotype	AA	Aa	aa
survival value	α	β	γ
frequencies	p^2	$2pq$	q^2

The last row gives the expected frequencies of the three genotypes if mating is at random.

If generations do not overlap, then the change in p in the course of one generation is

$$\Delta p = \frac{\alpha p^2 + \beta pq}{\alpha p^2 + 2\beta pq + \gamma q^2} - p = \frac{\frac{\alpha}{\beta} p^2 + pq}{\frac{\alpha}{\beta} p^2 + 2pq + \frac{\gamma}{\beta} q^2} - p \qquad (1)$$

This difference equation is not well adapted for mathematical treatment. And therefore, even in the case where generations are distinguishable, it seems convenient to replace it by a differential equation. In cases where the ratios $\alpha : \beta, \gamma : \beta$ do not greatly differ from unity, the differential equation will represent the course of the selection process to a good degree of approximation.

There is however an additional reason, apart from mathematical convenience, which makes it preferable to represent the process by means of a differential equation. In Nature, as well as in most experimental arrangements, generations are often not separated in time; on the contrary, they may overlap to such an extent, that no distinction is possible, and even the calculation of an average generation time is very difficult. The selective process is therefore best regarded as being a continuous one, and hence is most naturally represented by means of a differential equation.

We shall therefore set up a differential equation which represents a certain continuous process. It is then assumed that the actual process of selection when the generations overlap approximates to this continuous one.

The form of the differential equation is suggested by equation (1) and is obtained by various formal transformations of (1). Let dt be an element of time. Then in equation (1) substitute

$$\frac{\alpha}{\beta} = 1 - A\,dt$$
$$\frac{\gamma}{\beta} = 1 - C\,dt \qquad (2)$$

and set dp in place of Δp. The numbers A and C are essentially differences of Malthusian parameters.

We obtain

$$dp = \frac{p^2(1 - A\,dt) + pq}{1 - Ap^2\,dt - Cq^2\,dt} - p$$

$$= \frac{pq(Cq - Ap)\,dt}{1 - Ap^2\,dt - Cq^2\,dt}$$

which, on rejecting powers of dt higher than the first, becomes

$$\frac{dp}{dt} = pq(Cq - Ap) \tag{3}$$

On integration of (3) we obtain

$$\frac{1}{C}\log p + \frac{1}{A}\log q - \frac{A+C}{AC}\log\left|\frac{C}{A+C} - p\right| = t + \text{const.} \tag{4}$$

At equilibrium $dp = 0$, and the equilibrium gene frequencies p_E and q_E will be:

$$p_E = \frac{C}{A+C}, \quad q_E = 1 - p_E = \frac{A}{A+C}$$

We can therefore write equation (4) as follows:

$$\frac{1}{p_E}\log p + \frac{1}{q_E}\log q - \frac{1}{p_E q_E}\log|p_E - p| = (A+C)(t + \text{const.}) \tag{5}$$

Equation (5) was suggested by Professor Fisher as providing a transformation of gene frequencies which would give a function linear with time, if the equilibrium frequencies p_E and q_E are known.

Let us call

$$Y = \frac{1}{p_E}\log p + \frac{1}{q_E}\log q - \frac{1}{p_E q_E}\log|p_E - p| \tag{6}$$

the transformed gene frequencies, and

$$Y = a + bt \tag{7}$$

the linear function connecting Y and time of selection. The slope of the linear function, b, which is equal to $(A + C)$, in conjunction with the equilibrium frequency p_E will allow the calculation of the survival values A and C. The position of the straight line (7) given by a, will depend on the initial gene frequency, being equal to Y_0 (for time $t = 0$). The slope is independent of the initial conditions; therefore the problem of equality of $A + C$ in independent experiments, started with different initial frequencies of the same genes, will be a problem of parallelism of the corresponding straight lines after transformation (6). The goodness of fit of the straight line to the transformed gene frequencies will provide a test of the hypothesis that the selective process can actually be described, in a satisfactory way, by means of the differential equation given above.

THE PROCESS OF FITTING

The problem of estimating the parameters A and C, and the other

related problems mentioned before, are essentially centered on the process of fitting the straight line (7) to experimental data. Table I gives values of Y as a function of p and p_E. When p_E is exactly known, the observed gene frequencies will be transformed into Y by interpolation of the table, and the Y's will be plotted against time. Best estimation of a and b in equation (7) involves, however, weighting of the Y values and, as the weight will depend on the expected values of Y, only a process of successive approximation will in practice give best estimates for a and b.

TABLE I

TABLE OF THE VALUES OF THE FUNCTION:

$$Y = \frac{1}{p_E} \log p + \frac{1}{q_E} \log q - \frac{1}{p_E q_E} \log (p_E - p)$$

p \ p_E	0.50	0.45	0.40	0.35	0.30	0.25	0.20	0.15	0.10	0.05			
0.01	−6.38	−6.93	−7.61	−8.43	−9.47	−10.82	−12.66	−15.29	−19.31	−24.35			
0.05	−2.90	−3.05	−3.20	−3.35	−3.46	−3.47	−3.19	−1.97	+3.27			+25.52	0.95
0.10	−1.15	−1.07	−0.92	−0.65	−0.16	+0.77	+2.75	+8.02			+16.91	+12.86	0.90
0.15	+0.08	+0.35	+0.76	+1.40	+2.48	+4.48	+9.03			+14.13	+10.36	+8.80	0.85
0.20	+1.15	+1.62	+2.31	+3.40	+5.28	+9.24			+12.50	+9.24	+7.52	+6.59	0.80
0.25	+2.20	+2.90	+3.96	+5.72	+9.23			+11.43	+8.48	+6.90	+5.85	+5.23	0.75
0.30	+3.32	+4.34	+5.99	+9.18			+10.69	+7.93	+6.43	+5.45	+4.73	+4.28	0.70
0.35	+4.23	+6.19	+9.14			+10.15	+7.51	+6.07	+5.12	+4.43	+3.90	+3.56	0.65
0.40	+6.36	+9.14			+9.76	+7.18	+5.77	+4.84	+4.16	+3.65	+3.24	+2.97	0.60
0.45	+9.19			+9.49	+6.91	+5.52	+4.59	+3.92	+3.42	+3.02	+2.69	+2.47	0.55
			+9.30	+6.71	+5.29	+4.36	+3.70	+3.19	+2.80	+2.47	+2.22	+2.04	0.50
		+9.19	+6.52	+5.08	+4.14	+3.47	+2.97	+2.57	+2.26	+2.01	+1.80	+1.66	0.45
		+6.36	+4.86	+3.90	+3.22	+2.72	+2.33	+2.03	+1.78	+1.58	+1.40	+1.28	0.40
		+4.63	+3.64	+2.95	+2.45	+2.06	+1.76	+1.52	+1.33	+1.17	+1.03	+0.94	0.35
		+3.32	+2.62	+2.12	+1.74	+1.45	+1.23	+1.04	+0.89	+0.77	+0.67	+0.60	0.30
		+2.20	+1.70	+1.34	+1.07	+0.86	+0.70	+0.57	+0.46	+0.37	+0.30	+0.24	0.25
		+1.15	+0.82	+0.58	+0.40	+0.26	+0.15	+0.07	+0.00	−0.05	−0.10	−0.13	0.20
		+0.08	−0.11	−0.24	−0.34	−0.41	−0.45	−0.49	−0.52	−0.54	−0.55	−0.55	0.15
		−1.15	−1.19	−1.21	−1.22	−1.21	−1.19	−1.18	−1.16	−1.13	−1.11	−1.10	0.10
		−2.90	−2.76	−2.63	−2.51	−2.40	−2.30	−2.20	−2.12	−2.04	−1.96	−2.91	0.05
		−6.38	−5.91	−5.50	−5.15	−4.85	−4.57	−4.33	−4.12	−3.92	−3.75	−3.63	0.01
		0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	0.99	p \ p_E

In the more general case, however, the equilibrium frequencies are not known, or at least are not exactly known, and therefore must be estimated from the data. The same is true of the initial gene frequency, p_0, which will be known only within a sampling error. Therefore, the estimation of A and C will be obtained through the estimation of the three parameters a, b, of equation (7), and p_E, from the data.

Let us call θ_i any of the three parameters; choose p so that $p_E > p$; call n the number of observations on which each gene frequency is

determined; call p the observed, and P the expected (similarly $q = 1 - p$ and $Q = 1 - P$) gene frequencies. The maximum likelihood equations are

$$\frac{\partial L}{\partial \theta_i} = \sum \frac{n(p - P)}{PQ} \frac{\partial P}{\partial \theta_i} \tag{8}$$

there being one such equation for each of the three parameters. No direct solution of this system of equations is available, and therefore a method of successive approximation will be used, as is usual in other applications of maximum likelihood to related problems (see for instance, Probit Analysis, Finney 1947).

Equations (8) will be expanded according to Taylor-Maclaurin series, with consideration of first orders only, giving

$$\frac{\partial L}{\partial \theta_i} = \delta\theta_1 \frac{\partial \frac{\partial L}{\partial \theta_i}}{\partial \theta_1} + \delta\theta_2 \frac{\partial \frac{\partial L}{\partial \theta_i}}{\partial \theta_2} + \delta\theta_3 \frac{\partial \frac{\partial L}{\partial \theta_i}}{\partial \theta_3} \tag{9}$$

Trial values of the three parameters will be inserted in equations (9) and from this, the adjustments for each parameter, $\delta\theta_i$, will be calculated. The whole process can be repeated, using the adjusted values of the three parameters, and such cycles of computations will be repeated until no further adjustment is necessary (the adjustments being small in comparison of their standard error). All computations are greatly simplified by taking, in the second derivatives of L, $p = P$; in this case the system of equations will become, writing it in extenso,

$$\begin{cases} \sum \frac{n(p - P)}{PQ} \frac{\partial P}{\partial \theta_1} = \delta\theta_1 \sum \frac{n}{PQ} \left(\frac{\partial P}{\partial \theta_1}\right)^2 + \delta\theta_2 \sum \frac{n}{PQ} \frac{\partial P}{\partial \theta_1} \frac{\partial P}{\partial \theta_2} \\ \qquad + \delta\theta_3 \sum \frac{n}{PQ} \frac{\partial P}{\partial \theta_1} \frac{\partial P}{\partial \theta_3} \\ \sum \frac{n(p - P)}{PQ} \frac{\partial P}{\partial \theta_2} = \delta\theta_1 \sum \frac{n}{PQ} \frac{\partial P}{\partial \theta_1} \frac{\partial P}{\partial \theta_2} + \delta\theta_2 \sum \frac{n}{PQ} \left(\frac{\partial P}{\partial \theta_2}\right)^2 \\ \qquad + \delta\theta_3 \sum \frac{n}{PQ} \frac{\partial P}{\partial \theta_2} \frac{\partial P}{\partial \theta_3} \\ \sum \frac{n(p - P)}{PQ} \frac{\partial P}{\partial \theta_3} = \delta\theta_1 \sum \frac{n}{PQ} \frac{\partial P}{\partial \theta_1} \frac{\partial P}{\partial \theta_3} + \delta\theta_2 \sum \frac{n}{PQ} \frac{\partial P}{\partial \theta_2} \frac{\partial P}{\partial \theta_3} \\ \qquad + \delta\theta_3 \sum \frac{n}{PQ} \left(\frac{\partial P}{\partial \theta_3}\right)^2 \end{cases} \tag{10}$$

To obtain the three derivatives $\partial P/\partial\theta$, which, in the present instance, are $\partial P/\partial p_E$, $\partial P/\partial a$, $\partial P/\partial b$, let us consider the function

$$\psi \equiv q_E \log P + p_E \log Q - \log (p_E - P) - p_E q_E Y = 0 \quad (11)$$

We shall then have

$$\frac{\partial P}{\partial a} = \frac{-\partial\psi/\partial a}{\partial\psi/\partial P} = PQ(p_E - P)$$

and similarly

$$\frac{\partial P}{\partial b} = PQ(p_E - P)t\,, \quad \frac{\partial P}{\partial p_E} = -PQ(p_E - P)z$$

where z stands for

$$z = -\frac{1}{p_E q_E}\left[\log Q/P - \frac{1}{p_E - P} - (q_E - p_E)Y\right]$$

Inserting these values into equations (10), and setting

$$s = (p - P)(p_E - P)$$

$$w = PQ(p_E - P)^2$$

we shall obtain the following equations:

$$\left\{\begin{aligned} \textstyle\sum ns &= \delta a \textstyle\sum nw + \delta b \textstyle\sum nwt + \delta p_E \textstyle\sum nwz \\ \textstyle\sum nst &= \delta a \textstyle\sum nwt + \delta b \textstyle\sum nwt^2 + \delta p_E \textstyle\sum nwtz \\ \textstyle\sum nsz &= \delta a \textstyle\sum nwz + \delta b \textstyle\sum nwtz + \delta p_E \textstyle\sum nwz^2 \end{aligned}\right. \quad (12)$$

This method can be considered as a simple extension to the case of more than one parameter of the method of scoring, proposed for the analysis of linkage data by R. A. Fisher (1947), and which can be considered as the most direct form of solution of maximum likelihood equations by successive approximation. The quantities on the left in eq. (12) are then *scores* S_θ , in respect of the three parameters a, b, p_E ; the matrix of the coefficients of the adjustments is the matrix of information I, in respect of the estimates of the three parameters, or the variance matrix in respect of the scores. System (12), in matrix notation will be

$$\underset{\sim}{S_\theta} = \underset{\sim}{\delta} I$$

I being the information matrix and $\underset{\sim}{\delta}$ the vector of the adjustments δa, δb, δp_E . The adjustments are obtained from

$$\underset{\sim}{\delta} = \underset{\sim}{S}_{\theta} I^{-1}$$

and their variances from the elements of the principal diagonal in I^{-1}.

APPLICATION

In an experiment of artificial selection on Drosophila pseudoobscura, Dobzhansky established populations with various initial gene frequencies of certain chromosome arrangements, which, for our purposes, can be considered as single genes, and observed that an equilibrium was reached

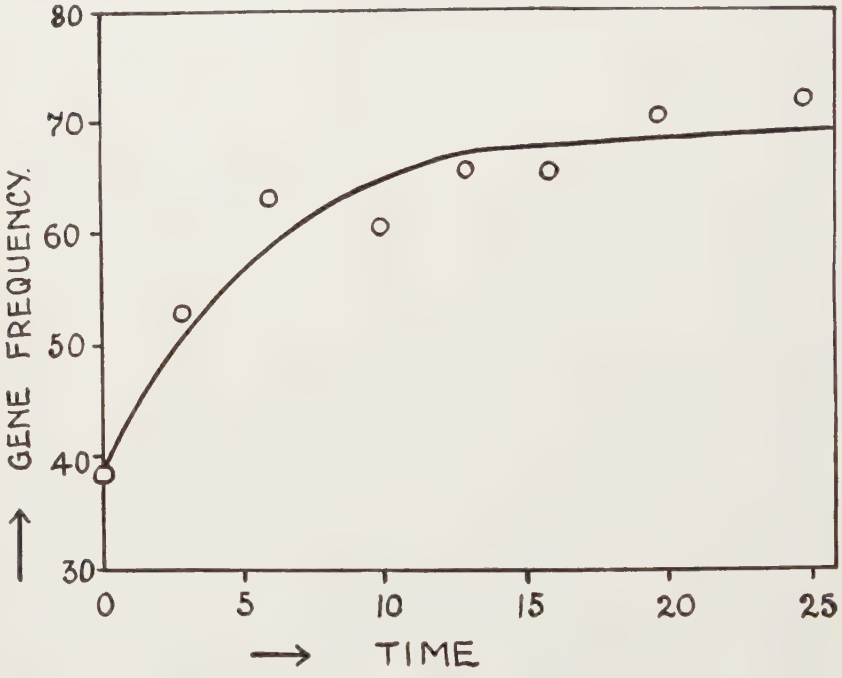


FIGURE I

after a certain number of generations. In one such experiment started with about one-third of chromosomes of the Standard type, and the rest of the Chiricahua type, and finished after nine months when about three fourths of the chromosomes were of the standard type, samples of about 150 individuals were taken and their chromosomes examined at various times after the mixture of the two original types. Frequencies observed at given times (time units of 10 days) are given in the second column of table II. Numbers of chromosomes counted are given as n in the third column. The same data are shown graphically in fig. 1.

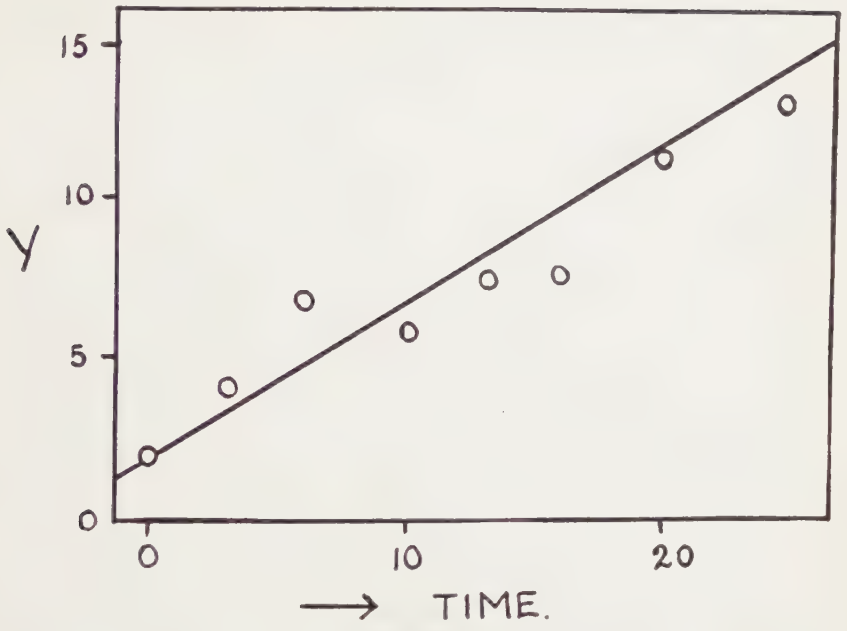


FIGURE II

The equilibrium frequency is not known; from an inspection of fig. 1, assuming that equilibrium had not yet been fully reached at time when observation ceased, a first trial value of p_E was taken at $p_E = 0.75$, and hence values of the transformation (6) corresponding to the observed gene frequencies were interpolated from table I and plotted in fig. 2 against time. These values are indicated as y in the fourth column of table II.

In fig. 2, a straight line was fitted by eye to the data, being

$$Y = 2.13 + 0.48t$$

First trial values will be indicated with the suffix 1. They are

$$a_1 = 2.13$$

$$b_1 = 0.48$$

$$p_{E_1} = 0.75$$

Expected values, calculated from the straight line fitted graphically, are indicated with Y in the fifth column of table II, and from Y, expected values P are interpolated by means of table I (or an extension of it).

TABLE II

FIRST CYCLE OF COMPUTATIONS ON DOBZHANSKY'S DATA

t	p	n	y	Y	P	nw	ns	z	nwz	nwt
0	.383	1278	2.1	2.13	.386	40.13	−1.396	+ 6.56	263.2528	0
3	.530	300	4.2	3.57	.491	5.04	+3.030	+ 10.93	55.0872	15.12
6	.633	300	6.8	5.01	.569	2.41	+3.475	+ 17.33	41.7653	14.46
10	.603	300	5.9	6.93	.635	0.920	−1.104	+ 30.88	28.4096	9.20
13	.653	300	7.6	8.37	.667	0.459	−0.349	+ 45.65	20.9533	5.967
16	.653	300	7.6	9.81	.689	0.239	−0.659	+ 65.44	15.6402	3.824
20	.704	250	11.5	11.73	.708	0.0912	−0.0420	+100.37	9.1537	1.824
25	.720	300	13.6	14.13	.722	0.0472	−0.0168	+157.81	7.4486	1.180
						49.3364	+2.9378		441.7107	51.575

By means of the formulas given in the second section, from p, P and Y and from $p_E = 0.75$ the values of s, w and z are calculated, and hence the last five columns of table II are obtained.

From these, the following other sums of squares and products are easily calculated:

$$S\,nwt^2 = 428.85; \qquad S\,nwtz = 1591.88; \qquad S\,nwz^2 = 8004.36$$

$$S\,nst = +2.559; \qquad S\,nsz = -15.8332$$

With such values, the system of the adjustment equations is formed:

scores information matrix × adjustments

$$+\ 2.938 = 49.34\ \delta a + 51.57\ \delta b + 441.71\ \delta p_E$$

$$+\ 2.559 = 51.57\ \delta a + 428.55\ \delta b + 1591.88\ \delta p_E$$

$$-15.833 = 441.71\ \delta a + 1591.88\ \delta b + 8004.36\ \delta p_E$$

The inverse of the matrix of information is

$$\begin{Bmatrix} +0.0755123 & +0.0244006 & -0.0090198 \\ +0.0244006 & +0.0167924 & -0.0046861 \\ -0.0090198 & -0.0046861 & +0.0015546 \end{Bmatrix}$$

and multiplying this by the matrix of scores, one secures the adjustments to the trial values of the three parameters

$$\delta a_1 = 0.0755123 \times 2.938 + 0.0244006 \times 2.559 + 0.0090198 \times 15.833$$

$$= 0.4271 \pm 0.275$$

$$\delta b_1 = 0.1889 \pm 0.130$$

$$\delta p_{E_1} = -0.06311 \pm 0.0394$$

The standard errors of the adjustments are the square roots of the corresponding elements of the principal diagonal in the inverted matrix; thus, the standard error of δa is $\sqrt{0.0755123} = 0.275$. The adjusted values of the three parameters, after this first cycle, will then be

$$a_2 = 2.56, \qquad b_2 = 0.67, \qquad p_{E_2} = 0.687$$

The adjustments are greater than their standard errors, and therefore it seems worthwhile repeating the computations, taking as estimates of the three parameters the adjusted values given above. Having performed this second cycle, the following adjustments were found:

$$\delta a_2 = -0.0222 \pm .255$$

$$\delta b_2 = +0.0557 \pm .167$$

$$\delta p_{E_2} = +0.0122 \pm .023$$

The adjustments are now well below their standard error; in normal cases, therefore, no further refinement would be necessary. In this case, however, a third cycle was carried out, and for reasons which will be apparent later, the inverse of the matrix of information is given in full below.

Third cycle. Inverse of the matrix of information — Scores

$$\left\{\begin{matrix} +0.0592353 & +0.0178369 & -0.0035900 \\ +0.0178369 & +0.0320025 & -0.0031734 \\ -0.0035900 & -0.0031734 & +0.0004880 \end{matrix}\right\} \quad \begin{matrix} -0.3908 \\ +1.1092 \\ -0.0825 \end{matrix}$$

The solutions for the adjustments are:

$$\delta a_3 = -0.00307 \pm .243$$

$$\delta b_3 = +0.02879 \pm .179$$

$$\delta p_{E_3} = -0.00216 \pm .022$$

and the final estimates of the three parameters are:

$$a_4 = 2.54, \qquad b_4 = 0.76, \qquad p_{E_4} = 0.697$$

In the final stage, the goodness of fit can be tested in the usual way by means of χ^2, comparing expected and observed gene frequencies. This has been here done at all stages, and the results are given below. χ^2 was calculated with three figure accuracy only for the last fit. There are, in all, eight observations, and three parameters have been estimated from the data, so there remain five degrees of freedom. The probability corresponding to each χ^2 is given in the bottom line.

	Graphical fitting	Analytical fitting		
		after 1st cycle	after 2nd cycle	after 3rd cycle
χ^2	10.32	9.43	7.18	7.036
P	$<.10$ $>.05$	$<.10$ $>.05$	$<.30$ $>.20$	$<.30$ $>.20$

One further point is the calculation of the values A and C, which are the ones that really matter from the genetical point of view. Since

$$p_E = \frac{C}{A + C}, \qquad q_E = \frac{A}{A + C}, \qquad b = A + C$$

we shall have

$$A = bq_E, \qquad C = bp_E$$

To calculate the standard errors of A and C, an approximate formula is available, and we shall need some of the values contained in the inverted information matrix. Calling V_b, V_{p_E} and W_{bp_E} the elements of the inverse information matrix of places c_{22}, c_{33}, c_{23}, we shall have:

$$V(A) = q_E^2 V_b + b^2 V_{p_E} - 2bq_E W_{bp_E}$$

$$V(C) = p_E^2 V_b + b^2 V_{p_E} + 2bp_E W_{bp_E}$$

Taking these values from the matrix obtained in the last cycle of computations (where $V_b = 0.0320025$; $V_{pE} = 0.0004880$; $W_{bpE} = -0.0031734$) we obtain numerically:

$$A = 0.230 \pm 0.063$$

$$C = 0.530 \pm 0.119$$

The values of A and C are expressed in the time scale used in the calculations, which is here a time unit of 10 days approximately. The solid line in fig. 1 represents the theoretical curve they calculated.

DISCUSSION

The method here developed allows the estimation of the parameters involved in a selection process, in which the heterozygous genotype has an advantage over both homozygotes, and therefore the frequency of the two alleles in competition comes to an equilibrium value, which depends on the relative survival values of the homozygous genotypes. When estimates of these parameters have been obtained, tests of goodness of fit become possible. It is however to be noticed that one necessary condition for the application, in the present problem, of the method of maximum likelihood and of the usual tests of goodness of fit, may not be fulfilled, namely the condition of the reciprocal independence of the observations. A selective process is a stochastic process, and a gene frequency at a given time depends on gene frequencies at earlier generations unless the population is infinite. While such difficulty need not worry us for most populations of animals and plants in Nature, it seems possible that it may need to be taken into consideration for artificial populations, which are usually of small size. The total effect will be of a greater variance than that due to sampling only, i.e. a bad fit—which might however arise through a multitude of other causes. In the example considered above, the deviation from the theoretical course of selection is not significant, and therefore there is no detectable evidence of drift or other cause of departure. Where drift is suspected, the only method of treatment so far known is that which has been proposed by Fisher and Ford for the analysis of a natural population of Panaxia dominula.

One consideration which applies to drift, however, is that when a small sample of the population is observed each time, the sampling variance will be many times greater than the variance of gene frequencies due to the finite size of the population. The smaller the ratio of the size of the sample to the size of the total population, the better the approximation to the condition of independence necessary for an unbiased application of the maximum likelihood, and of tests of goodness of fit. In the

example of which use has been made here, the size of the sample was 300, and the size of the population twenty times larger, a ratio that seems in fact to provide the conditions for the application of the test.

This investigation was started with the aim of analyzing experimental data by Prof. A. Buzzati-Traverso; the analysis of such data will, however, appear as an appendix to Buzzati's paper. The whole investigation would not have been possible without the generous advice and help of Prof. R. A. Fisher. My gratitude is also due to Dr. A. R. G. Owen for revision of the manuscript.

REFERENCES

Finney, D. J. 1947. *Probit analysis.* Cambridge University Press.

Fisher, R. A. 1922. On the dominance ratio. *Proc. R. Soc. Edinb. 42:* 321–341.

Fisher, R. A. 1946. A system of scoring linkage data, with special reference to the pied factor in mice. *Amer. Nat. 80:* 497–592.

Fisher, R. A. and E. B. Ford. 1947. The spread of a gene in natural conditions in a colony of the moth Panaxia dominula L. *Heredity 1:* 143–174.

Haldane, J. B. S. 1926. A mathematical theory of natural and artificial selection. Part III. *Proc. Camb. Phil. Soc. 23:* 363–372.

Wright, S. and Th. Dobzhansky. 1946. Genetics of natural populations. *Genetics, 31:* 125–156.

RECENT APPLICATIONS OF BIOMETRICAL METHODS IN GENETICS

(3) SCORES FOR THE ESTIMATION OF PARAMETERS

D. J. Finney

Lecturer in the Design and Analysis of Scientific Experiment, University of Oxford

The biometric statistician should not rest content with developing sound methods for the statistical analysis of biological data: he should also put these methods into the form in which they can most easily be applied by a biologist with no extensive knowledge of mathematics. He will thereby frequently ensure that the labour of statistical analysis forms only a small part of the total effort in a particular research project. Experimenters often argue their preference for a non-efficient method of estimation or analysis on the grounds that it is more easily understood and performed. This may be a false economy, for a rearrangement of the calculations and the provision of suitable tables may much simplify the work and make fully efficient, or at least very highly efficient, methods available with the minimum of labour.

In no branch of biology is the need for efficient statistical methods more evident than in genetics. The collection of genetical data may entail the breeding and recording of plants or animals for many years, or the patient search for informative human pedigrees. Yet records collected at great cost are often subjected to methods of analysis, chosen without regard to statistical efficiency, methods which may fail to extract as much as 50% of the information available.

A very convenient computing scheme for the estimation of genetic parameters, such as gene frequencies and recombination frequencies, is a scoring system which assigns a numerical score to each individual or each family in the records and bases the estimate upon an average score for all the data. Such a method can be very quickly applied, especially when tables of scores have been prepared.

Suppose that θ is a single unknown parameter to be estimated, say the frequency of recombinations between two loci. If we have data covering, say, two generations, all possible families may be classified

according to *patterns*, such that the relative frequencies of different patterns are independent of θ. For example, all records of two parents of specified phenotypes with a particular number of offspring would constitute one pattern, all records of a particular number of sibs for which no parents have been classified would constitute another, and so on. Within any one pattern, the probability of obtaining for the records individuals of particular phenotypes would be a function of θ, say $P(\theta)$: using $S(\ )$ for summation over all families of the same pattern

$$S(P) = 1.$$

Take θ_0 as an approximate estimate of θ, obtained by any rapid method (by inspection, or by rough analysis of all or part of the data) and write

$$\theta = \theta_0 + \delta,$$

where δ is an adjustment to θ_0. Then to the first order in δ

$$P(\theta) = P(\theta_0) + \delta \left.\frac{dP}{d\theta}\right|_{\theta=\theta_0}$$

$$= P_0 + P_0'\delta \qquad \text{say}$$

and

$$S(P_0) = 1,$$

$$S(P_0') = 0.$$

Let z be a *score* calculated from the observed phenotype frequencies in a family: z may be any function of the several frequencies, subject to the obvious restriction that its average value for all families of the same pattern must be dependent on θ. Now

$$E(z) = S(Pz)$$

$$= S(P_0z) + \delta S(P_0'z).$$

Provided that $S(P_0'z)$ is not zero, an average value of z from observations can be used in association with $S(P_0z)$ and $S(P_0'z)$ to give an estimate of δ: this is then added to θ_0 to give a revised estimate of θ. For convenience in handling, a score which leads directly to the revised estimate has advantages: write

$$y = \left\{\theta_0 - \frac{S(P_0z)}{S(P_0'z)}\right\} + \frac{z}{S(P_0'z)},$$

then

$$E(y) = \theta_0 + \delta$$

$$= \theta.$$

Consequently, simple averaging of y over all families of the same pattern gives an estimate of θ directly. Moreover, ignoring terms in δ,

$$V(y) = \frac{S(P_0 z^2) - \{S(P_0 z)\}^2}{\{S(P_0' z)\}^2}.$$

Hence, if for any one pattern of family we write

$$W = 1/V(y),$$

and then combine the scores from all the records into the weighted mean

$$\bar{y} = \frac{\sum Wy}{\sum W}$$

(where $\sum$ denotes summation over all data), $\bar{y}$ will be the most precise average value of y that can be formed and will have expectation θ. Therefore, $\bar{y}$ may be taken as θ_1, a revised estimate of θ.

If y and W happen to be independent of θ_0, this is the end of the calculation: $\bar{y}$ is the best estimate of θ that can be based upon the scores y. More usually, however, both y and W will be functions of θ_0, and iterative scoring must be practised. Values of y and W are obtained for an initial θ_0; these give $\bar{y}$, which is taken as θ_1; new sets of values of y and W are then based on θ_1, and the new $\bar{y}$ is taken as θ_2. This continues until two successive θ's are sufficiently close together for their difference to be ignored. The latest value is then taken as the estimate, θ_e, and

$$S.E.(\theta_e) = \frac{1}{\sqrt{\sum W}}.$$

The procedure sounds laborious, but in fact is simple and rapid if tables of y and W as functions of θ_0 have been prepared for different patterns of family.

The estimate reached depends upon the form of z adopted in each pattern of family. Indeed, the method is a general expression of many existing methods for the estimation of genetic parameters, which consist merely of equating some arbitrary function of phenotype frequencies to its expectation. A criterion is now needed for deciding between rival scoring systems. As is well known, an asymptotically efficient estimate

of θ, that is to say an estimate which in large samples has the least possible variance, may be obtained by the method of maximum likelihood; this requires that

$$\sum (\log P)$$

be maximized for the data. This may be accomplished by scoring with

$$Z = \frac{d}{d\theta} (\log P) \Big|_{\theta = \theta_0}$$

$$= \frac{P_0'}{P_0} .$$

The corresponding y score is

$$Y = \theta_0 + \frac{P_0'}{WP_0} ,$$

with

$$W = S\left(\frac{P_0'^2}{P_0}\right).$$

The weight W is a well-known expression for the total information provided by a single family of particular pattern if $\theta = \theta_0$, thus demonstrating the efficiency of the estimate based on Y. It is convenient to tabulate Y and

$$\Lambda = WY = W\theta_0 + \frac{P_0'}{P_0}$$

for different values of θ_0, so that the revised estimate may easily be obtained as

$$\theta_1 = \frac{\sum \Lambda}{\sum W} .$$

Often Λ and W are both linear functions of the phenotype frequencies amongst the progeny of a family, so that for specified parental types all family sizes may be covered by tabulation of weight and score per individual. The iteration goes as before.

The efficiency of any other z-score may be assessed by comparing the information extracted by it with the total amount available:

$$\text{Eff.}(z) = \frac{\{S(P_0'z)\}^2}{S\left(\frac{P_0'^2}{P_0}\right)[S(P_0z^2) - \{S(P_0z)\}^2]} .$$

This formula assists a decision as to whether a score other than the maximum likelihood Y involves a sacrifice of information unreasonably great in comparison with the saving of labour. It is sometimes possible to find a scoring system different from Y and more easily calculated which is nevertheless of full efficiency, but more usually other scores will sacrifice at least a little of the available information.

Similar methods can be applied when two or more parameters must be estimated simultaneously, though the calculations are then necessarily more complicated. The maximum likelihood estimation of the parameters of a normal tolerance distribution by means of the probit transformation is an example outside the field of genetics (Finney, 1947; 1949a).

Tables for certain applications of the method to gene frequency and recombination frequency estimation have been given elsewhere (Finney, 1948a, b; 1949b), together with illustrations of their use.

REFERENCES

Finney, D. J. (1947). The principles of biological assay. *Supplement to the Journal of the Royal Statistical Society, 9,* 46–91.

Finney, D. J. (1948a). The estimation of gene frequency from family records. I. Factors without dominance. *Heredity, 2,* 199–218.

Finney, D. J. (1948b). The estimation of gene frequency from family records. II. Factors with dominance. *Heredity, 2,* 369–389.

Finney, D. J. (1949a). The estimation of the parameters of tolerance distributions. *Biometrika, 36,* 239–256.

Finney, D. J. (1949b). The estimation of the frequency of recombinations. I. Matings of known phase. *Journal of Genetics, 49,* 159–176.

(These papers contain bibliographies more complete than can be given here and show how the ideas were first developed by R. A. Fisher and others.)

DISCUSSION ON RECENT APPLICATIONS OF BIOMETRICAL METHODS IN GENETICS

(1) EXPERIMENTAL TECHNIQUES IN PLANT IMPROVEMENT

F. Yates

W. G. Cochran. Some time ago I presented the ideas in Dr. Yates' paper to a group of plant breeders in the United States. Using a number of plots that appeared typical and typical estimated values for the genetic and environmental variances, I showed that the optimum number of varieties would allow only one or two replications and that the F value had only a small chance of being significant. The audience was somewhat shocked, apparently because my results seemed to contradict the usual advice of the statistician that several replications must be used and that F tests must be made. This suggests that rather patient explanation will be needed in order to make Dr. Yates' results clear and acceptable to plant breeders.

With regard to selection in several stages, some unpublished work on two-stage selection has been conducted in the United States. The work suggests, as a speculation, that the optimum is fairly flat, i.e. that a considerable deviation from the optimum procedure does not greatly decrease the genetic advance.

R. A. Fisher. The selectionist requires either significant, or at least subsignificant evidence of genetic variability at the first stage of quantitative testing. His program for the next 3–5 years may be decided on the strength of this evidence. So I can understand that American plant breeders should be a little shocked at the suggestion that the z test should be ignored at the first stage of selection.

G. Pompilj. Je voudrai développer quelques considérations critiques tendant à éclaircir la signification logique du problème du "testing".

Je crois qu'on doit distinguer ce qu'on voudrait obtenir et ce qu'on peut obtenir. Ce qu'on voudrait obtenir est très clair. Nous voudrions obtenir des procédés, tout à fait objectifs et impersonnels, pour classifier, selon certains caractères, les différentes lignes ou variétés qu'on a obtenues.

Mais les données à notre disposition sont elles en général suffisantes pour ce but? Je crois que non. En effet le problème que nous sommes posé, comme monsieur Gini l'a observé il y a plusieurs années, demande, pour être résolu, deux ordres de connaissances; à savoir: la distribution *a priori* de toutes les variétés en tenant compte des caractères prefissées et la distribution des mêmes variétés apres les expériences. Or, en général, nous ne connaissons pas la distribution *a priori;* il nous manque donc un des deux éléments nécessaires pour resoudre le problème.

(2) THE ANALYSIS OF SELECTION CURVES

Luigi L. Cavalli

J. B. S. Haldane. It is extremely gratifying that Dr. Cavalli has obtained such good agreement with a simple hypothesis. Insofar as there were no deviations from linearity, the data are consistent with the following hypotheses.

1. Mating was at random, the influence of inbreeding and assortative mating being negligible.

2. Selection was Darwinean, that is to say it proceeded as if the three genotypes survived to maturity in different proportions, so that only two parameters were needed. Often selection may depend on the interaction of two genotypes, as when a foetus heterozygous for Rh sensitivity has a decidedly different expectation of life according as its mother is or is not Rh-sensitive. If the q different matings, e.g $AA\,♀ \times AA\,♂$, $Aa\,♀ \times aa\,♂$ have different fertilities, we should reach a five parameter equation.

3. Selection was of constant intensity, independent of the composition of the population, although this might, for example, have altered the culture medium.

L. L. Cavalli. I entirely agree with Professor Haldane about the simplicity of the hypothesis considered. However, I believe that as a general policy introduction of new parameters may be dispensed with until goodness of fit with old parameters is satisfactory, and as far, of course, as old parameters make sense from a biological point of view.

(3) SCORES FOR THE ESTIMATION OF PARAMETERS

D. J. Finney

M. J. R. Healy. When more than one parameter is to be estimated, several cycles of iteration are more frequently required. In this case the process leads to simultaneous equations. During the iteration, it will save computation to adjust only the right hand side of these equations, exact values of the left hand side being only required at the final stage to obtain variances and covariances.

G. Pompilj. Je voudrai donner un avertissement qui se rattache à ce que j'ai déja dit à propos de la première communication. Il s'agit d'une faute de logique qui, partant de la statistique, va infecter les autres branches de la science et même de la biométrie.

Je parle des théories de signification et d'évaluation des paramètres, dont l'inconsistence logique a été démonstrée par monsieur Gini il y a dix ans. La chose est trop longue à développer ici. Je me limiterai donc à vous recommander, si ça n'a pas trop de présomption, de vous méfier toujours de tous les "tests" de signification et de toutes les théories qui promettent d'atteindre à l'évaluation des paramètres, car on peut démontrer que pour obtenir des résultats corrects il faut connaitre des éléments qui en général ne sont pas à notre connaissance.

Other participants in the discussion included F. Chodat and F. Bernstein.

INDUSTRIAL APPLICATIONS OF BIOMETRY

BIOMETRIC METHODS IN THE CHEMICAL INDUSTRY

O. L. Davies

Biological Laboratories, Imperial Chemical Industries Ltd., Hexagon House, Blackley, Manchester

Abstract

In the Chemical Industry a great deal of research is directed to the improvement of yield and quality of chemical products. This research usually consists of assessing experimentally either in the laboratory or on the plant, the effect of various changes in the process conditions, for instance, changes in the temperature and time of reaction, concentration and variety of the reactants. Several of these factors may be examined in the same experiment, and factorial arrangements lend themselves particularly well to this kind of research. In these experiments the observations are made sequentially, that is, one after another, and this gives a large measure of flexibility to the experimental design, because it is possible to modify the design during the course of the experiment in the light of the results obtained. Partial factorial designs in sequence are particularly adaptable to this type of experimental work, and two experiments relating to the problem of improving the persistence of penicillin in the body are described to illustrate these principles.

Simple methods are introduced for the construction of partial factorial designs and these are illustrated by an investigation on the large scale manufacture of penicillin, in which a relatively large number of factors were investigated. In investigations on the plant scale it is usually necessary to allow for time trends and this may be done by using the well-known principles of confounding, with "blocks" spaced in time. Typical partial factorial designs for the 2^n and 3^n systems are given.

In experimental work in industry we frequently know from past experience the magnitude of the experimental error. This information may be used to assess the significance of the effects in a partial factorial experiment when there are insufficient degrees of freedom in the experiment itself to obtain a reliable estimate of the experimental error. Another use is in estimating the size of a projected experiment and the considerations involved here are the control of the risks of errors of the first and second kind. The levels decided for the risks of these two types of errors are dictated by the economic consequence of making the errors of the first and second kind.

DISCUSSION FOLLOWING BIOMETRIC METHODS IN THE CHEMICAL INDUSTRY

O. L. Davies

Besse B. Day. Dr. Davies has furnished us with an invaluable guide to the practical use of partial factorials. Industrial statisticians are deeply indebted to biometricians for many modern statistical techniques which are equally applicable to engineering problems but often with a different emphasis. Thus the prime interest may be in the error term itself—how reproducible are the results?

Among the applications of interest in both biological and industrial fields is a war-time statistical development in ordnance testing which may have application to research for determining the 50% lethal dose. By this method of sensitivity testing the height from which a specimen of explosive is dropped is changed by a fixed increment after each specimen has been tested, increased if the specimen did not explode and decreased if it did explode. This ensures many results near the 50% point from which the mean height for 50% and its standard deviation may be estimated. A tool with biological and industrial applications is spectrographic analysis. The evaluation of its reliability and of conversion factors or formulas are statistical problems which require experimental designs, regression analysis, etc. Bacteriological counts have an industrial parallel in estimating solid contaminants in used lubricating oils from crankcases. In general the problems in industrial experimentation requiring statistical tools are (1) instrumentation, more practically calibration, (2) development and standardization of test methods, (3) specifications—setting of tolerances, (4) improvement of a product and (5) development of new products.

E. C. Fieller. Dr. Davies' paper was to be admired not only for its adaptations of modern experimental design to the needs of his industry, but also for the wise restraint with which these applications were made, and the care taken to ensure beforehand that the conditions under which fractional designs might be validly used were in fact justified.

As an example of the difficulties that might arise without these preliminary assurances, an experiment in the design of radio valves has been described, which started off on a highly fractional basis but had eventually to be made a complete factorial experiment.

Many speakers at this and previous sessions have found difficulty in distinguishing between Statistics and Biometry but perhaps the answer is suggested by A. C. Bacharach's remark that "Every chemical experiment involves some living organism, even if it is only the chemist."

A. Hald. I should like to stress two points. First, the need for statistical investigations of the testing technique. Very often the laboratory error is considerably larger than that shown by the usual duplicate analyses, because the conditions of the laboratory may vary in time and produce a trend in the measurements. This trend does not become apparent from duplicate analyses and is often more important than the errors of measurement proper. Therefore, it is necessary to keep the laboratory errors under statistical control.

The second point to be noted is the sequential nature of industrial experiments and the corresponding time trend. The fitting of trends with polynomials or with a moving average seem to me well suited to take this time factor into account and might be considered as useful supplements to the usual analysis of variance. Especially as industrial experiments are often very costly, it may pay to use regression analysis instead of the analysis of variance.

H. Åstrand. The discussion of Dr. Davies' paper makes me doubt that "biometrics" describes it adequately. Would it not be better to speak of "statistical methods", in spite of the fact that to a large extent they are developed by biometricians?

It seems that the field for applying statistical methods in industry is immense. I feel sure that many more statisticians from industry will take part in future meetings of our Society and with that expectation I close the session.

NEWS AND NOTES

The Survey Research Center of the University of Michigan will hold its Third Annual Summer Institute in Survey Research Techniques from July 24 to August 18, 1950. The following courses will be offered: Introduction to Survey Research, Survey Research Methods, Sampling Methods in Survey Research (Introductory), Sampling Methods in Survey Research (Advanced), Mathematics of Sampling, Statistical Methods in Survey Research, and Techniques of Scaling. In addition, the introductory courses will be given from June 26 to July 21. This will permit students who are attending the full eight-week summer session of the University (June 26 to August 16) to register for the introductory courses during the first four weeks. It is expected that this special session will attract men and women employed in business or governmental research or other statistical work and university instructors and graduate students with a particular interest in this area of social science research. All courses are offered for graduate credit and students must be admitted by the Graduate School . . . Yale University's Committee on Statistics offered a special course in applied statistics during recent months. The following speakers with their topics took part in this course. **J. W. Tukey** (Princeton) Order statistics, signed ranks, **W. J. Youden** (Bureau of Standards) Measurements in the physical sciences, **D. F. Votaw, Jr.** (Yale) Matrix theory for statisticians, **P. J. Rulon** (Harvard) Discriminant function in psychology and education, **H. Hotelling** (North Carolina) Multivariate analysis, **C. I. Bliss** (Yale) Discriminant function in biological assay, **F. Mosteller** (Harvard) Statistical problems in social psychology, **T. W. Anderson** (Columbia) Use of stochastic equation models in economics, **R. A. Fisher**(Cambridge) Enumeration and recombination in genetics, **W. G. Cochran** (Johns Hopkins) Estimation of components of variance and their use, **W. Allen Wallis** (Chicago) Prediction of future observations, sequential analysis, statistics of the Kinsey Report, **S. S. Wilks** (Princeton) Industrial statistics, **H. F. Dorn** (National Institute of Health) Statistical problems in the measurement of morbidity, and **Donald Mainland** (Dalhousie) Statistics in clinical medicine. During his stay at Yale, each visiting lecturer was available for a limited number of individual or group conferences. . . . **L. Otis Emik** is in the Commissioned Corps of the United States Public Health Service. He is Scientist with the Statistical Branch, Communicable Disease Center at Atlanta. His duties are to develop a program of Veterinary Statistics. Both census and research programs

are being organized, dealing with those diseases which affect both man and animals. Mr. Emik before 1950 was with the United States Sheep Experiment Station and Western Sheep Breeding Laboratory at Dubois, Idaho. . . . **E. L. LeClerg** spent the month of January at Mayaguez, Puerto Rico, giving daily lectures on experimental design to the technical staff of the Federal Experiment Station. He is Research Coordinator in Field Crops for the Agricultural Research Administration of the United States Department of Agriculture at Washington, D. C. . . . **F. E. Satterthwaite** is quality control engineer of the Plastics Division in the Chemical Department of the General Electric Company in Pittsfield, Massachusetts, according to an announcement by F. W. Warner, engineering manager of the division. Mr. Satterthwaite has been quality control engineer for the product service division of the Company in Bridgeport, Connecticut. . . . **B. Schneider,** formerly with the University of West Virginia, Morgantown, is now with the Department of Animal Husbandry, The State College of Washington, Pullman.